

# Regional Trends in UN Patent Data for the Period 2010 and 2020

Roger M. Rosewall, D.Eng.

2023-02-01

## Introduction

This case study analyzes patent data from the World Intellectual Property Office (WIPO) curated by the United Nations (UN). Patent activity is considered a proxy measure of research and development (R&D). The analysis presented here investigates some of the trends in patent activity between the years 2010 and 2020. The analysis relies on ISO 3166, *Codes for the representation of names of countries and their subdivisions*, a standard published by the International Organization for Standardization (ISO) that defines codes for the names of countries, dependent territories, special areas of geographical interest, and their principal subdivisions (e.g., provinces or states). The standard employs a code of letters and numbers to represent the name of a given geographical area in order to save time and energy when describing the area, as well as to reduce the risk of description errors.

ISO 3166 recognizes the following geographical regions:

- Africa
- Americas
- Asia
- Europe
- Oceania

The regions are divided into geographical sub-regions, and individual countries are associated with a sub-region. The analysis provides answers to the following questions.

## Research Questions

The following questions are posed to understand the trends in R&D throughout the world during the period 2010 to 2020.

### Patents granted in 2010

- Q1. Which region had the largest number of patents granted in 2010?
- A1. Asia, the countries of Asia had the largest total number of patents granted in 2010: 466,526, approximately 54.9% of all patents granted in 2010.
- Q2. Which sub-region of that region had the largest number of patents granted in 2010?
- A2. Eastern Asia, the countries of Eastern Asia had the largest total of patents granted in Asia in 2010: 438,541, approximately 94% of all patents granted in Asia in 2010.
- Q3. Which country in that sub-region had the largest number of patents granted in 2010?
- A3. Japan had the largest number of patents granted in Eastern Asia in 2010: 222,693, approximately 50.78% of all patents granted to countries in Eastern Asia in 2010.

## Patents in force in 2010

- Q4. Which region had the largest number of patents in force in 2010?
- Q4. Europe, the countries of Europe had the largest total number of patents in force in 2010: 2,998,246, approximately 36.24% of all patents in force in 2010.
- Q5. Which sub-region of that region had the largest number of patents in force in 2010?
- A5. Western Europe, the countries of Europe had the largest total number of patents in force in 2010: 1,496,897 patents, 49.93% of all patents in force in Europe in 2010.
- Q6. Which country in that sub-region had the largest number of patents in force in 2010?
- A6. Germany had the largest number of patents granted in Western Europe in 2010: 514,046, approximately 34.34% of all patents in force in Western Europe in 2010.

## Patents granted in 2020

- Q7. Which region had largest number of patents granted in 2020?
- A7. Asia, the countries of Asia had the largest total number of patents granted in 2020: 927,612, approximately 63.7% of all patents granted in 2020.
- Q8. Which sub-region of that region had the largest number of patents granted in 2020?
- A8. Eastern Asia, the countries included in Eastern Asia had the largest total number of patents granted in 2020: 856,303, 92.31% of all patents granted in Asia in 2020.
- Q9. Which country in that sub-region had the largest number of patents granted in 2020?
- A9. China had the largest number patents granted in Eastern Asia in 2020: 530,127, approximately 61.91% of all patents patents granted to countries in Eastern Asia.

## Patents in force in 2020

- Q10. Which region had the largest number of patents in force in 2020?
- A10. Asia, the countries of Asia had the largest total number of patents in force in 2020: 6,721,553, approximately 42.46% of all patents in force in 2020.
- Q11. Which sub-region of that region had the largest number of patents in force in 2020?
- A11. Eastern Asia, the countries of Eastern Asia had the largest total number of patents in force in 2020: 6,249,001, 92.97% of all patents in force in Asia.
- Q12. Which country had the largest number of patents in force in 2020?
- A12. China had the largest number of patents in force in Eastern Asia in 2020: 3,057,844, approximately 48.93% of all patents in force in the countries in Eastern Asia.

## Data sources

The analysis presented in this case study is based on two datasets. The primary dataset is WIPO patent data curated by the UN. The patent dataset is available at <http://data.un.org> and the file name is *SYB65\_264\_202209\_Patents.csv*. The second data set provides International Standardization Organization (ISO) information about geographic regions and sub-regions associated with individual countries. That dataset is available at <https://github.com/luke/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv>, and the file name is *ISO-3166-Countries-with-Regional-Codes*.

## Method

The two datasets were imported into Rstudio, transformed, cleaned, and merged into a single dataframe. Using that dataframe four subsets (dataframes) were created, corresponding to the four groups of questions posed in the Introduction: Patents Granted in 2010, Patents in Force in 2010, Patents Granted in 2020, and Patentes in Force in 2020. Those four subsets were analyzed to answer the questions posed.

## Prepare for Analysis

The following tasks are required to prepare the data for analysis:

- Setup the environment
- Import and prepare the datasets
- Merge the two datasets
- Extract relevant data into new dataframes
- Format the created dataframes

### Setup the environment

Install Packages and load libraries

```
## The URL ensures the packages are installed from the Comprehensive R Archive Network (CRAN)
install.packages("tidyverse", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
## /var/folders/36/m4yk12555hlc6gphrs1cf1c0000gn/T//Rtmpzq1942/downloaded_packages
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
install.packages("ggplot2", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
## /var/folders/36/m4yk12555hlc6gphrs1cf1c0000gn/T//Rtmpzq1942/downloaded_packages
library(ggplot2)
```

### Import and prepare the two datasets

The two data sets must be imported and then organized to ensure their data formats, types, and structures enable merging the datasets.

## Import and prepare the UN patent data

Import the dataset curated by the UN. The patent dataset is available at <http://data.un.org>. The name is SYB65\_264\_202209\_Patents.csv. The following code chunk reads the csv file into an R dataframe *named* `un_patent_data`

```
un_patent_data <- read.csv("~/Desktop/SYB65_264_202209_Patents.csv",  
                           header=FALSE)
```

The following code chunk presents a tibble identifying the patent attributes included in the doataframe.

```
tibble(un_patent_data)
```

```
## # A tibble: 1,741 x 7  
##   V1          V2          V3    V4    V5    V6    V7  
##   <chr>      <chr>      <chr> <chr> <chr> <chr> <chr>  
## 1 T31       "Patents"      ""    ""    ""    ""    ""  
## 2 Region/Country/Area ""            "Yea~ "Ser~ "Val~ "Foo~ "Sou~  
## 3 1         "Total, all countries or a~ "198~ "Gra~ "397~ ""    "Wor~  
## 4 1         "Total, all countries or a~ "199~ "Gra~ "430~ ""    "Wor~  
## 5 1         "Total, all countries or a~ "200~ "Gra~ "633~ ""    "Wor~  
## 6 1         "Total, all countries or a~ "201~ "Gra~ "914~ ""    "Wor~  
## 7 1         "Total, all countries or a~ "201~ "Gra~ "1,4~ ""    "Wor~  
## 8 1         "Total, all countries or a~ "201~ "Gra~ "1,5~ ""    "Wor~  
## 9 1         "Total, all countries or a~ "202~ "Gra~ "1,5~ ""    "Wor~  
## 10 2        "Africa"      "200~ "Gra~ "4,8~ ""    "Wor~  
## # ... with 1,731 more rows
```

**Rename the columns** The following code chunk renames the columns with more descriptive text than the default: V1, V2, V3, etc.

```
# Set column names  
colnames(un_patent_data) = c("country_code", "region", "year", "series", "number", "footnotes", "source")  
tibble(un_patent_data)
```

```
## # A tibble: 1,741 x 7  
##   country_code region          year series number footn~1 source  
##   <chr>      <chr>      <chr> <chr> <chr> <chr> <chr>  
## 1 T31       "Patents"      ""    ""    ""    ""    ""  
## 2 Region/Country/Area ""            "Yea~ "Seri~ "Valu~ "Footn~ "Sour~  
## 3 1         "Total, all countries~ "198~ "Gran~ "397,~ ""    "Worl~  
## 4 1         "Total, all countries~ "199~ "Gran~ "430,~ ""    "Worl~  
## 5 1         "Total, all countries~ "200~ "Gran~ "633,~ ""    "Worl~  
## 6 1         "Total, all countries~ "201~ "Gran~ "914,~ ""    "Worl~  
## 7 1         "Total, all countries~ "201~ "Gran~ "1,42~ ""    "Worl~  
## 8 1         "Total, all countries~ "201~ "Gran~ "1,50~ ""    "Worl~  
## 9 1         "Total, all countries~ "202~ "Gran~ "1,59~ ""    "Worl~  
## 10 2        "Africa"      "200~ "Gran~ "4,80~ ""    "Worl~  
## # ... with 1,731 more rows, and abbreviated variable name 1: footnotes
```

**Delete unneeded rows** The following code chunk deletes the first two rows of `un_patent_data` dataframe, since they serve no useful purpose for this analysis.

```
un_patent_data <- un_patent_data[-c(1,2),]  
tibble(un_patent_data)
```

```
## # A tibble: 1,739 x 7
##   country_code region          year series number footn~1 source
##   <chr>         <chr>         <chr> <chr> <chr> <chr> <chr>
## 1 1           Total, all countries or areas 1985 Grant~ 397,5~ ""      World~
## 2 1           Total, all countries or areas 1995 Grant~ 430,5~ ""      World~
## 3 1           Total, all countries or areas 2005 Grant~ 633,1~ ""      World~
## 4 1           Total, all countries or areas 2010 Grant~ 914,2~ ""      World~
## 5 1           Total, all countries or areas 2018 Grant~ 1,423~ ""      World~
## 6 1           Total, all countries or areas 2019 Grant~ 1,501~ ""      World~
## 7 1           Total, all countries or areas 2020 Grant~ 1,592~ ""      World~
## 8 2           Africa                2005 Grant~ 4,800  ""      World~
## 9 2           Africa                2010 Grant~ 9,000  ""      World~
## 10 2          Africa                2018 Grant~ 8,700  ""      World~
## # ... with 1,729 more rows, and abbreviated variable name 1: footnotes
```

**Delete unneeded columns** The following code chunk deletes both the *footnotes* column and the *source* column. The “footnotes” column includes a single item of information: “For statistical purposes, the data for China do not include those for the Hong Kong Special Administrative Region (Hong Kong SAR), Macao Special Administrative Region (Macao SAR) and Taiwan Province of China.” Taking note of this information, the column is not needed. The *source* column is the same for all rows, acknowledging WIPO as the data source, therefore this column is not needed.

```
un_patent_data <- un_patent_data[,-c(6,7)]
tibble(un_patent_data)
```

```
## # A tibble: 1,739 x 5
##   country_code region          year series          number
##   <chr>         <chr>         <chr> <chr>          <chr>
## 1 1           Total, all countries or areas 1985 Grants of patents (n~ 397,5~
## 2 1           Total, all countries or areas 1995 Grants of patents (n~ 430,5~
## 3 1           Total, all countries or areas 2005 Grants of patents (n~ 633,1~
## 4 1           Total, all countries or areas 2010 Grants of patents (n~ 914,2~
## 5 1           Total, all countries or areas 2018 Grants of patents (n~ 1,423~
## 6 1           Total, all countries or areas 2019 Grants of patents (n~ 1,501~
## 7 1           Total, all countries or areas 2020 Grants of patents (n~ 1,592~
## 8 2           Africa                2005 Grants of patents (n~ 4,800
## 9 2           Africa                2010 Grants of patents (n~ 9,000
## 10 2          Africa                2018 Grants of patents (n~ 8,700
## # ... with 1,729 more rows
```

**Extract the records relevant to the time period of interest** The following code chunk removes from the dataframe all rows where the year is not equal to either 2010 or 2020, since those are the years considered in this analysis, and stores the result in a new dataframe, *un\_patent\_data\_2010\_and\_2020*.

```
un_patent_data_2010_and_2020 <- subset(un_patent_data, year=='2010' | year== '2020' )
tibble(un_patent_data_2010_and_2020)
```

```
## # A tibble: 576 x 5
##   country_code region          year series          number
##   <chr>         <chr>         <chr> <chr>          <chr>
## 1 1           Total, all countries or areas 2010 Grants of patents (n~ 914,2~
## 2 1           Total, all countries or areas 2020 Grants of patents (n~ 1,592~
## 3 2           Africa                2010 Grants of patents (n~ 9,000
## 4 2           Africa                2020 Grants of patents (n~ 7,000
```

```
## 5 419 Latin America & the Caribbean 2010 Grants of patents (n~ 17,200
## 6 419 Latin America & the Caribbean 2020 Grants of patents (n~ 36,100
## 7 142 Asia 2010 Grants of patents (n~ 469,6~
## 8 142 Asia 2020 Grants of patents (n~ 924,5~
## 9 150 Europe 2010 Grants of patents (n~ 160,8~
## 10 150 Europe 2020 Grants of patents (n~ 231,5~
## # ... with 566 more rows
```

**Identify the current data types** The following code chunk identifies the current data type of each column in the dataframe.

```
print(sapply(un_patent_data_2010_and_2020, class))
```

```
## country_code      region      year      series      number
## "character" "character" "character" "character" "character"
```

**Convert charater data types to numeric** The following code chunk converts the *country\_code* to numeric.

```
un_patent_data_2010_and_2020$country_code = as.numeric(as.character(un_patent_data_2010_and_2020$country_code))
```

The following code chunk converts the *year* to numeric.

```
un_patent_data_2010_and_2020$year = as.numeric(as.character(un_patent_data_2010_and_2020$year))
```

The following code chunk removes commas from the characters in the *number* column. The commas must be removed before converting to numeric, otherwise they will not convert correctly.

```
un_patent_data_2010_and_2020$number<-gsub(",", "", as.character(un_patent_data_2010_and_2020$number))
```

The following code chunk converts *number* to numeric.

```
un_patent_data_2010_and_2020$number = as.numeric(as.character(un_patent_data_2010_and_2020$number))
tibble(un_patent_data_2010_and_2020)
```

```
## # A tibble: 576 x 5
##   country_code region      year series      number
##   <dbl> <chr>      <dbl> <chr>      <dbl>
## 1         1 Total, all countries or areas 2010 Grants of patents (n~ 9.14e5
## 2         1 Total, all countries or areas 2020 Grants of patents (n~ 1.59e6
## 3         2 Africa 2010 Grants of patents (n~ 9 e3
## 4         2 Africa 2020 Grants of patents (n~ 7 e3
## 5       419 Latin America & the Caribbean 2010 Grants of patents (n~ 1.72e4
## 6       419 Latin America & the Caribbean 2020 Grants of patents (n~ 3.61e4
## 7       142 Asia 2010 Grants of patents (n~ 4.70e5
## 8       142 Asia 2020 Grants of patents (n~ 9.24e5
## 9       150 Europe 2010 Grants of patents (n~ 1.61e5
## 10      150 Europe 2020 Grants of patents (n~ 2.32e5
## # ... with 566 more rows
```

## Import and prepare the ISO geographical data

Import the dataset ISO-3166-Countries-with-Regional-Codes, <https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv>. The following code chunk reads the dataset into the dataframe *ISO\_1366\_Countries\_Regions*.

```
ISO_1366_Countries_Regions <- read.csv("~/Desktop/ISO-3166-Countries-with-Regional-Codes.csv",
                                         header=FALSE)
tibble(ISO_1366_Countries_Regions)
```

```
## # A tibble: 250 x 11
##   V1      V2      V3      V4      V5      V6      V7      V8      V9      V10     V11
##   <chr>    <chr>    <chr>    <chr>    <chr>    <chr>    <chr>    <chr>    <chr>    <chr>    <chr>
## 1 name      alpha-2 alpha~ coun~ iso_~ "reg~ "sub~ "int~ "reg~ "sub~ "int~
## 2 Afghanistan AF      AFG      004      ISO ~ "Asi~ "Sou~ ""      "142" "034" ""
## 3 Åland Islands AX      ALA      248      ISO ~ "Eur~ "Nor~ ""      "150" "154" ""
## 4 Albania    AL      ALB      008      ISO ~ "Eur~ "Sou~ ""      "150" "039" ""
## 5 Algeria    DZ      DZA      012      ISO ~ "Afr~ "Nor~ ""      "002" "015" ""
## 6 American Samoa AS      ASM      016      ISO ~ "Oce~ "Pol~ ""      "009" "061" ""
## 7 Andorra    AD      AND      020      ISO ~ "Eur~ "Sou~ ""      "150" "039" ""
## 8 Angola     AO      AGO      024      ISO ~ "Afr~ "Sub~ "Mid~ "002" "202" "017"
## 9 Anguilla   AI      AIA      660      ISO ~ "Ame~ "Lat~ "Car~ "019" "419" "029"
## 10 Antarctica AQ      ATA      010      ISO ~ ""      ""      ""      ""      ""
## # ... with 240 more rows
```

**Remove unneeded columns** The following code chunk removes columns from the dataframe that are not relevant for the analysis: *alpha-2*, *iso\_3166-2*, *intermediate-region*, and *intermediate-region-code*.

```
ISO_1366_Countries_Regions <- ISO_1366_Countries_Regions[,-c(2,5,8,11)]
tibble(ISO_1366_Countries_Regions)
```

```
## # A tibble: 250 x 7
##   V1      V3      V4      V6      V7      V9      V10
##   <chr>    <chr>    <chr>    <chr>    <chr>    <chr>    <chr>
## 1 name      alpha-3 country-code "region" "sub-region" "reg~ "sub~
## 2 Afghanistan AFG      004      "Asia"    "Southern Asia" "142" "034"
## 3 Åland Islands ALA      248      "Europe"  "Northern Europe" "150" "154"
## 4 Albania    ALB      008      "Europe"  "Southern Europe" "150" "039"
## 5 Algeria    DZA      012      "Africa"  "Northern Africa" "002" "015"
## 6 American Samoa ASM      016      "Oceania" "Polynesia" "009" "061"
## 7 Andorra    AND      020      "Europe"  "Southern Europe" "150" "039"
## 8 Angola     AGO      024      "Africa"  "Sub-Saharan Afri~ "002" "202"
## 9 Anguilla   AIA      660      "Americas" "Latin America an~ "019" "419"
## 10 Antarctica ATA      010      ""        ""        ""        ""
## # ... with 240 more rows
```

**Rename the columns** The following code chunk provides the columns with more descriptive text than the default: V1, V2, V3, etc.

```
colnames(ISO_1366_Countries_Regions) = c("country", "alpha-3", "country_code", "region", "sub_region",
tibble(ISO_1366_Countries_Regions)
```

```
## # A tibble: 250 x 7
##   country `alpha-3` country_code region    sub_region regio-1 sub_r-2
##   <chr>    <chr>    <chr>    <chr>    <chr>    <chr>    <chr>
## 1 name      alpha-3 country-code "region" "sub-region" "regio~ "sub-r~
## 2 Afghanistan AFG      004      "Asia"    "Southern A~ "142" "034"
## 3 Åland Islands ALA      248      "Europe"  "Northern E~ "150" "154"
## 4 Albania    ALB      008      "Europe"  "Southern E~ "150" "039"
## 5 Algeria    DZA      012      "Africa"  "Northern A~ "002" "015"
```

```
## 6 American Samoa ASM      016      "Oceania" "Polynesia" "009" "061"
## 7 Andorra      AND      020      "Europe"  "Southern E~ "150" "039"
## 8 Angola      AGO      024      "Africa"  "Sub-Sahara~ "002" "202"
## 9 Anguilla     AIA      660      "Americas" "Latin Amer~ "019" "419"
## 10 Antarctica  ATA      010      ""        ""        ""      ""
## # ... with 240 more rows, and abbreviated variable names 1: region_code,
## # 2: sub_region_code
```

**Remove unneeded rows** The following code chunk removes the first row, since it serves no purpose.

```
ISO_1366_Countries_Regions <- ISO_1366_Countries_Regions[-1,]
tibble(ISO_1366_Countries_Regions)
```

```
## # A tibble: 249 x 7
##   country      `alpha-3` country_code region      sub_r~1 regio~2 sub_r~3
##   <chr>      <chr>      <chr>      <chr>      <chr>  <chr>  <chr>
## 1 Afghanistan AFG      004      "Asia"      "South~ "142" "034"
## 2 Åland Islands ALA      248      "Europe"    "North~ "150" "154"
## 3 Albania      ALB      008      "Europe"    "South~ "150" "039"
## 4 Algeria      DZA      012      "Africa"    "North~ "002" "015"
## 5 American Samoa ASM      016      "Oceania"   "Polyn~ "009" "061"
## 6 Andorra      AND      020      "Europe"    "South~ "150" "039"
## 7 Angola      AGO      024      "Africa"    "Sub-S~ "002" "202"
## 8 Anguilla     AIA      660      "Americas"  "Latin~ "019" "419"
## 9 Antarctica  ATA      010      ""          ""        ""      ""
## 10 Antigua and Barbuda ATG      028      "Americas"  "Latin~ "019" "419"
## # ... with 239 more rows, and abbreviated variable names 1: sub_region,
## # 2: region_code, 3: sub_region_code
```

The following code chunk removes the 9th row (Antarctica) because the record is incomplete and it has no associated patents.

```
ISO_1366_Countries_Regions <- ISO_1366_Countries_Regions[-9,]
tibble(ISO_1366_Countries_Regions)
```

```
## # A tibble: 248 x 7
##   country      `alpha-3` country_code region      sub_reg~1 regio~2 sub_r~3
##   <chr>      <chr>      <chr>      <chr>      <chr>  <chr>  <chr>
## 1 Afghanistan AFG      004      Asia      Southern~ 142    034
## 2 Åland Islands ALA      248      Europe    Northern~ 150    154
## 3 Albania      ALB      008      Europe    Southern~ 150    039
## 4 Algeria      DZA      012      Africa    Northern~ 002    015
## 5 American Samoa ASM      016      Oceania   Polynesia 009    061
## 6 Andorra      AND      020      Europe    Southern~ 150    039
## 7 Angola      AGO      024      Africa    Sub-Saha~ 002    202
## 8 Anguilla     AIA      660      Americas  Latin Am~ 019    419
## 9 Antigua and Barbuda ATG      028      Americas  Latin Am~ 019    419
## 10 Argentina   ARG      032      Americas  Latin Am~ 019    419
## # ... with 238 more rows, and abbreviated variable names 1: sub_region,
## # 2: region_code, 3: sub_region_code
```

**Convert charater data types to numeric** The country\_code, region\_code, and sub\_region\_code are converted to numeric data types in the following code chunk.



```
## Convert character to numeric
ISO_1366_Countries_Regions$country_code = as.numeric(as.character(ISO_1366_Countries_Regions$country_code))
ISO_1366_Countries_Regions$region_code = as.numeric(as.character(ISO_1366_Countries_Regions$region_code))
ISO_1366_Countries_Regions$sub_region_code = as.numeric(as.character(ISO_1366_Countries_Regions$sub_region_code))
tibble(ISO_1366_Countries_Regions)
```

```
## # A tibble: 248 x 7
##   country      `alpha-3` country_code region  sub_reg-1 regio-2 sub_r-3
##   <chr>      <chr>      <dbl> <chr>    <chr>      <dbl>  <dbl>
## 1 Afghanistan AFG          4 Asia    Southern~  142    34
## 2 Åland Islands ALA        248 Europe Northern~  150   154
## 3 Albania     ALB          8 Europe Southern~  150    39
## 4 Algeria     DZA         12 Africa Northern~    2    15
## 5 American Samoa ASM         16 Oceania Polynesia    9    61
## 6 Andorra     AND         20 Europe Southern~  150    39
## 7 Angola      AGO         24 Africa Sub-Saha~    2   202
## 8 Anguilla    AIA        660 Americas Latin Am~   19   419
## 9 Antigua and Barbuda ATG         28 Americas Latin Am~   19   419
## 10 Argentina  ARG         32 Americas Latin Am~   19   419
## # ... with 238 more rows, and abbreviated variable names 1: sub_region,
## #    2: region_code, 3: sub_region_code
```

## Merge the two dataframes

The following code chunk merges the two dataframes, *un\_patent\_data\_2010\_and\_2020* and *ISO\_1366\_Countries\_Regions*, into a new dataframe *un\_patents\_with\_iso\_codes* using *country\_code* as the key variable. (The *country\_code* is used rather than *country* because country names can vary in spelling and capitalization, but the *country\_code* is standardized.)

```
un_patents_with_iso_codes <- merge(un_patent_data_2010_and_2020, ISO_1366_Countries_Regions, by="country_code")
tibble(un_patents_with_iso_codes)
```

```
## # A tibble: 564 x 11
##   country~1 regio~2 year series number country alpha~3 regio~4 sub_r~5 regio~6
##   <dbl> <chr> <dbl> <chr> <dbl> <chr> <chr> <chr> <chr> <dbl>
## 1      8 Albania 2010 Grant~  349 Albania ALB Europe Southe~  150
## 2      8 Albania 2020 Grant~    5 Albania ALB Europe Southe~  150
## 3      8 Albania 2010 Paten~  349 Albania ALB Europe Southe~  150
## 4      8 Albania 2020 Paten~ 5833 Albania ALB Europe Southe~  150
## 5     12 Algeria 2010 Resid~    2 Algeria DZA Africa Northe~    2
## 6     12 Algeria 2020 Resid~    4 Algeria DZA Africa Northe~    2
## 7     12 Algeria 2010 Grant~ 1076 Algeria DZA Africa Northe~    2
## 8     12 Algeria 2020 Grant~  421 Algeria DZA Africa Northe~    2
## 9     20 Andorra 2020 Grant~   13 Andorra AND Europe Southe~  150
## 10    20 Andorra 2020 Paten~   33 Andorra AND Europe Southe~  150
## # ... with 554 more rows, 1 more variable: sub_region_code <dbl>, and
## #    abbreviated variable names 1: country_code, 2: region.x, 3: `alpha-3`,
## #    4: region.y, 5: sub_region, 6: region_code
```

**Remove redundant columns** The following code chunk removes the column *region.x* and retains the *country* column from the ISO dataframe, because it is the standardized name.

```
un_patents_with_iso_codes <- un_patents_with_iso_codes[,-c(2)]
tibble(un_patents_with_iso_codes)
```

```
## # A tibble: 564 x 10
##   country~1 year series number country alpha~2 regio~3 sub_r~4 regio~5 sub_r~6
##   <dbl> <dbl> <chr>   <dbl> <chr>   <chr>   <chr>   <chr>   <dbl>   <dbl>
## 1      8  2010 Grant~    349 Albania ALB     Europe Southe~    150     39
## 2      8  2020 Grant~      5 Albania ALB     Europe Southe~    150     39
## 3      8  2010 Paten~    349 Albania ALB     Europe Southe~    150     39
## 4      8  2020 Paten~   5833 Albania ALB     Europe Southe~    150     39
## 5     12  2010 Resid~      2 Algeria DZA     Africa Northe~      2     15
## 6     12  2020 Resid~      4 Algeria DZA     Africa Northe~      2     15
## 7     12  2010 Grant~   1076 Algeria DZA     Africa Northe~      2     15
## 8     12  2020 Grant~    421 Algeria DZA     Africa Northe~      2     15
## 9     20  2020 Grant~     13 Andorra AND     Europe Southe~    150     39
## 10    20  2020 Paten~     33 Andorra AND     Europe Southe~    150     39
## # ... with 554 more rows, and abbreviated variable names 1: country_code,
## # 2: `alpha-3`, 3: region.y, 4: sub_region, 5: region_code,
## # 6: sub_region_code
```

**Reorder the columns** The following code chunk reorders the columns to improve understanding and readability.

```
un_patents_with_iso_codes <- un_patents_with_iso_codes[, c(1,5,6,2,4,3,7,8,9,10)]
tibble(un_patents_with_iso_codes)
```

```
## # A tibble: 564 x 10
##   country~1 country alpha~2 year number series regio~3 sub_r~4 regio~5 sub_r~6
##   <dbl> <chr>   <chr>   <dbl> <dbl> <chr>   <chr>   <chr>   <dbl>   <dbl>
## 1      8 Albania ALB     2010    349 Grant~ Europe Southe~    150     39
## 2      8 Albania ALB     2020      5 Grant~ Europe Southe~    150     39
## 3      8 Albania ALB     2010    349 Paten~ Europe Southe~    150     39
## 4      8 Albania ALB     2020   5833 Paten~ Europe Southe~    150     39
## 5     12 Algeria DZA     2010      2 Resid~ Africa Northe~      2     15
## 6     12 Algeria DZA     2020      4 Resid~ Africa Northe~      2     15
## 7     12 Algeria DZA     2010   1076 Grant~ Africa Northe~      2     15
## 8     12 Algeria DZA     2020    421 Grant~ Africa Northe~      2     15
## 9     20 Andorra AND     2020     13 Grant~ Europe Southe~    150     39
## 10    20 Andorra AND     2020     33 Paten~ Europe Southe~    150     39
## # ... with 554 more rows, and abbreviated variable names 1: country_code,
## # 2: `alpha-3`, 3: region.y, 4: sub_region, 5: region_code,
## # 6: sub_region_code
```

**Rename the columns** The following code chunk renames the columns to improve clarity.

```
colnames(un_patents_with_iso_codes) = c("country_code", "country", "iso_alpha_3", "year", "number_patent")
tibble(un_patents_with_iso_codes)
```

```
## # A tibble: 564 x 10
##   country~1 country iso_a~2 year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr>   <chr>   <dbl> <dbl> <chr>   <chr>   <chr>   <dbl>   <dbl>
## 1      8 Albania ALB     2010    349 Grant~ Europe Southe~    150     39
## 2      8 Albania ALB     2020      5 Grant~ Europe Southe~    150     39
## 3      8 Albania ALB     2010    349 Paten~ Europe Southe~    150     39
```

```
## 4      8 Albania ALB      2020    5833 Paten~ Europe Southe~    150    39
## 5     12 Algeria DZA      2010         2 Resid~ Africa Northe~        2    15
## 6     12 Algeria DZA      2020         4 Resid~ Africa Northe~        2    15
## 7     12 Algeria DZA      2010    1076 Grant~ Africa Northe~        2    15
## 8     12 Algeria DZA      2020     421 Grant~ Africa Northe~        2    15
## 9     20 Andorra AND      2020        13 Grant~ Europe Southe~    150    39
## 10    20 Andorra AND      2020        33 Paten~ Europe Southe~    150    39
## # ... with 554 more rows, and abbreviated variable names 1: country_code,
## # 2: iso_alpha_3, 3: number_patents, 4: sub_region, 5: region_code,
## # 6: sub_region_code
```

## Extract relevant data into new dataframes

This analysis is focused on patents granted and patents in force in 2010 and 2020. Four new dataframes are created to facilitate the analysis.

### Create dataframe for patents granted in 2010

The following code chunk creates the dataframe *un\_patents\_granted\_2010*.

```
## Patents granted in 2010
un_patents_granted_2010 <- subset(un_patents_with_iso_codes, year == 2010 & series == "Grants of patents
select=country_code:sub_region_code)
tibble(un_patents_granted_2010)

## # A tibble: 100 x 10
##   country~1 country iso_a~2   year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr>   <chr>   <dbl>   <dbl> <chr> <chr> <chr>    <dbl>   <dbl>
## 1      8 Albania ALB     2010     349 Grant~ Europe Southe~    150     39
## 2     12 Algeria DZA     2010    1076 Grant~ Africa Northe~        2     15
## 3     31 Azerba~ AZE     2010     126 Grant~ Asia Wester~    142    145
## 4     32 Argent~ ARG     2010    1366 Grant~ Ameri~ Latin ~     19   419
## 5     36 Austra~ AUS     2010   14557 Grant~ Ocean~ Austra~        9     53
## 6     40 Austria AUT     2010    1130 Grant~ Europe Wester~    150    155
## 7     50 Bangla~ BGD     2010      92 Grant~ Asia Southe~    142     34
## 8     51 Armenia ARM     2010     124 Grant~ Asia Wester~    142    145
## 9     56 Belgium BEL     2010     532 Grant~ Europe Wester~    150    155
## 10    70 Bosnia~ BIH     2010     173 Grant~ Europe Southe~    150     39
## # ... with 90 more rows, and abbreviated variable names 1: country_code,
## # 2: iso_alpha_3, 3: number_patents, 4: sub_region, 5: region_code,
## # 6: sub_region_code
```

### Create dataframe for patents granted in 2020

The following code chunk creates the dataframe *un\_patents\_granted\_2020*.

```
## Patents granted in 2020
un_patents_granted_2020 <- subset(un_patents_with_iso_codes, year == 2020 & series == "Grants of patents
select=country_code:sub_region_code)
tibble(un_patents_granted_2020)

## # A tibble: 122 x 10
##   country~1 country iso_a~2   year numbe~3 series region sub_r~4 regio~5 sub_r~6
```

```
##      <dbl> <chr>  <chr>    <dbl>    <dbl> <chr>  <chr>  <chr>    <dbl>    <dbl>
## 1         8 Albania ALB      2020        5 Grant~ Europe Southe~    150     39
## 2        12 Algeria DZA      2020     421 Grant~ Africa Northe~      2     15
## 3        20 Andorra AND      2020     13 Grant~ Europe Southe~    150     39
## 4        24 Angola  AGO      2020     33 Grant~ Africa Sub-Sa~      2    202
## 5        31 Azerba~ AZE      2020    110 Grant~ Asia  Wester~    142    145
## 6        32 Argent~ ARG      2020   2337 Grant~ Ameri~ Latin ~      19   419
## 7        36 Austra~ AUS      2020  17778 Grant~ Ocean~ Austra~      9     53
## 8        40 Austria AUT      2020   1058 Grant~ Europe Wester~    150    155
## 9        48 Bahrain BHR      2020     34 Grant~ Asia  Wester~    142    145
## 10       50 Bangla~ BGD      2020    140 Grant~ Asia  Southe~    142     34
## # ... with 112 more rows, and abbreviated variable names 1: country_code,
## #   2: iso_alpha_3, 3: number_patents, 4: sub_region, 5: region_code,
## #   6: sub_region_code
```

### Create dataframe for patents in force in 2010

The following code chunk creates the dataframe *un\_patents\_in\_force\_2010*.

```
## Patents in force in 2010
un_patents_in_force_2010 <- subset(un_patents_with_iso_codes, year == 2010 & series == "Patents in force"
select=country_code:sub_region_code)
tibble(un_patents_in_force_2010)
```

```
## # A tibble: 77 x 10
##   country~1 country iso_a~2  year numbe~3 series region sub_r~4 regio~5 sub_r~6
##       <dbl> <chr>  <chr>    <dbl>    <dbl> <chr>  <chr>  <chr>    <dbl>    <dbl>
## 1         8 Albania ALB      2010     349 Paten~ Europe Southe~    150     39
## 2        36 Austra~ AUS      2010   96293 Paten~ Ocean~ Austra~      9     53
## 3        40 Austria AUT      2010  102113 Paten~ Europe Wester~    150    155
## 4        51 Armenia ARM      2010     278 Paten~ Asia  Wester~    142    145
## 5        52 Barbad~ BRB      2010      57 Paten~ Ameri~ Latin ~      19   419
## 6        56 Belgium BEL      2010   89999 Paten~ Europe Wester~    150    155
## 7        70 Bosnia~ BIH      2010     716 Paten~ Europe Southe~    150     39
## 8        76 Brazil  BRA      2010   40022 Paten~ Ameri~ Latin ~      19   419
## 9       100 Bulgar~ BGR      2010    6812 Paten~ Europe Easter~    150    151
## 10      112 Belarus BLR      2010    4444 Paten~ Europe Easter~    150    151
## # ... with 67 more rows, and abbreviated variable names 1: country_code,
## #   2: iso_alpha_3, 3: number_patents, 4: sub_region, 5: region_code,
## #   6: sub_region_code
```

### Create dataframe for patents in force in 2020

The following code chunk creates the dataframe *un\_patents\_in\_force\_2020*.

```
## Patents in force in 2020
un_patents_in_force_2020 <- subset(un_patents_with_iso_codes, year == 2020 & series == "Patents in force"
select=country_code:sub_region_code)
tibble(un_patents_in_force_2020)
```

```
## # A tibble: 112 x 10
##   country~1 country iso_a~2  year numbe~3 series region sub_r~4 regio~5 sub_r~6
##       <dbl> <chr>  <chr>    <dbl>    <dbl> <chr>  <chr>  <chr>    <dbl>    <dbl>
## 1         8 Albania ALB      2020   5833 Paten~ Europe Southe~    150     39
```

```
## 2      20 Andorra AND      2020      33 Paten~ Europe Southe~      150      39
## 3      24 Angola  AGO      2020      85 Paten~ Africa Sub-Sa~      2      202
## 4      31 Azerba~ AZE      2020     252 Paten~ Asia Wester~      142     145
## 5      32 Argent~ ARG      2020    14550 Paten~ Ameri~ Latin ~      19     419
## 6      36 Austra~ AUS      2020   159304 Paten~ Ocean~ Austra~      9      53
## 7      40 Austria AUT      2020   159581 Paten~ Europe Wester~      150     155
## 8      48 Bahrain BHR      2020     131 Paten~ Asia Wester~      142     145
## 9      50 Bangla~ BGD      2020     1732 Paten~ Asia Southe~      142      34
## 10     51 Armenia ARM      2020     189 Paten~ Asia Wester~      142     145
## # ... with 102 more rows, and abbreviated variable names 1: country_code,
## # 2: iso_alpha_3, 3: number_patents, 4: sub_region, 5: region_code,
## # 6: sub_region_code
```

## Format the four created dataframes

The four dataframes used to answer the research questions require additional manipulation, specifically the addition of a column identifying percentages.

### Tabulate totals for the years 2010 and 2020

The following code chunk tabulates the annual totals for the periods of interest.

```
## total number of patents granted in 2010
total_granted_2010 = sum(un_patents_granted_2010$number_patents)
total_granted_2010

## [1] 849744

## total number of patents granted in 2020
total_granted_2020 = sum(un_patents_granted_2020$number_patents)
total_granted_2020

## [1] 1456299

## total number of patents in force in 2010
total_in_force_2010 = sum(un_patents_in_force_2010$number_patents)
total_in_force_2010

## [1] 8273884

## total number of patents in force in 2020
total_in_force_2020 = sum(un_patents_in_force_2020$number_patents)
total_in_force_2020

## [1] 15831644
```

### Calculate percentages

The following code chunk calculates percentage of patents granted in 2010 for each country as a share of the total number granted in 2010 and adds a new column *percent\_granted* to the dataframe.

```
## Percent of patents granted in 2010
un_patents_granted_2010$percent_granted <- (un_patents_granted_2010$number_patents / total_granted_2010)
tibble(un_patents_granted_2010)
```

```
## # A tibble: 100 x 11
##   country~1 country iso_a~2  year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      8 Albania ALB    2010    349 Grant~ Europe Southe~    150    39
## 2     12 Algeria DZA    2010   1076 Grant~ Africa Northe~     2    15
## 3     31 Azerba~ AZE    2010    126 Grant~ Asia Wester~    142   145
## 4     32 Argent~ ARG    2010   1366 Grant~ Ameri~ Latin ~    19   419
## 5     36 Austra~ AUS    2010  14557 Grant~ Ocean~ Austra~     9    53
## 6     40 Austria AUT    2010   1130 Grant~ Europe Wester~    150   155
## 7     50 Bangla~ BGD    2010     92 Grant~ Asia Southe~    142    34
## 8     51 Armenia ARM    2010    124 Grant~ Asia Wester~    142   145
## 9     56 Belgium BEL    2010    532 Grant~ Europe Wester~    150   155
## 10    70 Bosnia~ BIH    2010    173 Grant~ Europe Southe~    150    39
## # ... with 90 more rows, 1 more variable: percent_granted <dbl>, and
## # abbreviated variable names 1: country_code, 2: iso_alpha_3,
## # 3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
```

The following code chunk converts the percentage from scientific notation to decimal for readability.

```
un_patents_granted_2010$percent_granted <- round(un_patents_granted_2010$percent_granted, 3)
tibble(un_patents_granted_2010)
```

```
## # A tibble: 100 x 11
##   country~1 country iso_a~2  year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      8 Albania ALB    2010    349 Grant~ Europe Southe~    150    39
## 2     12 Algeria DZA    2010   1076 Grant~ Africa Northe~     2    15
## 3     31 Azerba~ AZE    2010    126 Grant~ Asia Wester~    142   145
## 4     32 Argent~ ARG    2010   1366 Grant~ Ameri~ Latin ~    19   419
## 5     36 Austra~ AUS    2010  14557 Grant~ Ocean~ Austra~     9    53
## 6     40 Austria AUT    2010   1130 Grant~ Europe Wester~    150   155
## 7     50 Bangla~ BGD    2010     92 Grant~ Asia Southe~    142    34
## 8     51 Armenia ARM    2010    124 Grant~ Asia Wester~    142   145
## 9     56 Belgium BEL    2010    532 Grant~ Europe Wester~    150   155
## 10    70 Bosnia~ BIH    2010    173 Grant~ Europe Southe~    150    39
## # ... with 90 more rows, 1 more variable: percent_granted <dbl>, and
## # abbreviated variable names 1: country_code, 2: iso_alpha_3,
## # 3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
```

The following code chunk calculates the percentage of patents granted in 2020 for each country as a share of the total number granted in 2020, converts it from scientific notation to decimal, and adds a new column *percent\_granted* to the dataframe.

```
## Percent of patents granted in 2020
un_patents_granted_2020$percent_granted <- (un_patents_granted_2020$number_patents / total_granted_2020)
un_patents_granted_2020$percent_granted <- round(un_patents_granted_2020$percent_granted, 3)
tibble(un_patents_granted_2020)
```

```
## # A tibble: 122 x 11
##   country~1 country iso_a~2  year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      8 Albania ALB    2020     5 Grant~ Europe Southe~    150    39
## 2     12 Algeria DZA    2020   421 Grant~ Africa Northe~     2    15
## 3     20 Andorra AND    2020    13 Grant~ Europe Southe~    150    39
## 4     24 Angola AGO    2020    33 Grant~ Africa Sub-Sa~     2   202
## 5     31 Azerba~ AZE    2020   110 Grant~ Asia Wester~    142   145
```

```
## 6      32 Argent~ ARG      2020      2337 Grant~ Ameri~ Latin ~      19      419
## 7      36 Austra~ AUS      2020     17778 Grant~ Ocean~ Austra~      9      53
## 8      40 Austria AUT      2020      1058 Grant~ Europe Wester~     150     155
## 9      48 Bahrain BHR      2020        34 Grant~ Asia  Wester~     142     145
## 10     50 Bangla~ BGD      2020       140 Grant~ Asia  Southe~     142      34
## # ... with 112 more rows, 1 more variable: percent_granted <dbl>, and
## #   abbreviated variable names 1: country_code, 2: iso_alpha_3,
## #   3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
```

The following code chunk calculates the percentage of patents in force in 2010 for each country as a share of the total number in force in 2010, converts it from scientific notation to decimal, and adds a new column *percent\_in\_force* to the dataframe.

```
## Percent of patents_in_force_2010
un_patents_in_force_2010$percent_in_force <- (un_patents_in_force_2010$number_patents / total_in_force_2010)
un_patents_in_force_2010$percent_in_force <- round(un_patents_in_force_2010$percent_in_force, 3)
tibble(un_patents_in_force_2010)
```

```
## # A tibble: 77 x 11
##   country~1 country iso_a~2  year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      8 Albania ALB      2010      349 Paten~ Europe Southe~     150      39
## 2     36 Austra~ AUS      2010    96293 Paten~ Ocean~ Austra~      9      53
## 3     40 Austria AUT      2010   102113 Paten~ Europe Wester~     150     155
## 4     51 Armenia ARM      2010      278 Paten~ Asia  Wester~     142     145
## 5     52 Barbad~ BRB      2010       57 Paten~ Ameri~ Latin ~      19     419
## 6     56 Belgium BEL      2010   89999 Paten~ Europe Wester~     150     155
## 7     70 Bosnia~ BIH      2010      716 Paten~ Europe Southe~     150      39
## 8     76 Brazil  BRA      2010   40022 Paten~ Ameri~ Latin ~      19     419
## 9    100 Bulgar~ BGR      2010    6812 Paten~ Europe Easter~     150     151
## 10   112 Belarus BLR      2010    4444 Paten~ Europe Easter~     150     151
## # ... with 67 more rows, 1 more variable: percent_in_force <dbl>, and
## #   abbreviated variable names 1: country_code, 2: iso_alpha_3,
## #   3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
```

The following code chunk calculates the percentage of patents in force in 2020 for each country as a share of the total number in force in 2020, converts it from scientific notation to decimal, and adds a new column *percent\_in\_force* to the dataframe.

```
## Percent of patents_in_force_2020
un_patents_in_force_2020$percent_in_force <- (un_patents_in_force_2020$number_patents / total_in_force_2020)
un_patents_in_force_2020$percent_in_force <- round(un_patents_in_force_2020$percent_in_force, 3)
tibble(un_patents_in_force_2020)
```

```
## # A tibble: 112 x 11
##   country~1 country iso_a~2  year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      8 Albania ALB      2020    5833 Paten~ Europe Southe~     150      39
## 2     20 Andorra AND      2020      33 Paten~ Europe Southe~     150      39
## 3     24 Angola  AGO      2020      85 Paten~ Africa Sub-Sa~      2     202
## 4     31 Azerba~ AZE      2020     252 Paten~ Asia  Wester~     142     145
## 5     32 Argent~ ARG      2020   14550 Paten~ Ameri~ Latin ~      19     419
## 6     36 Austra~ AUS      2020  159304 Paten~ Ocean~ Austra~      9      53
## 7     40 Austria AUT      2020  159581 Paten~ Europe Wester~     150     155
## 8     48 Bahrain BHR      2020     131 Paten~ Asia  Wester~     142     145
## 9     50 Bangla~ BGD      2020    1732 Paten~ Asia  Southe~     142      34
```



```
## 10      51 Armenia ARM      2020      189 Paten~ Asia  Wester~      142      145
## # ... with 102 more rows, 1 more variable: percent_in_force <dbl>, and
## # abbreviated variable names 1: country_code, 2: iso_alpha_3,
## # 3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
```

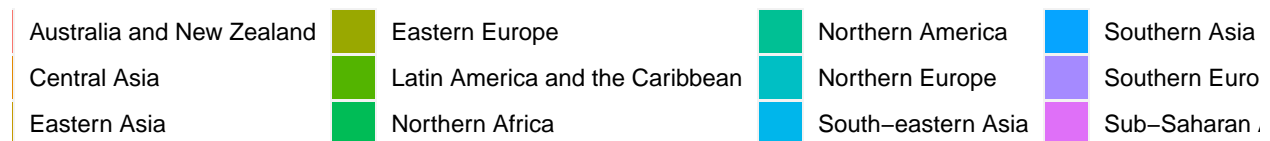
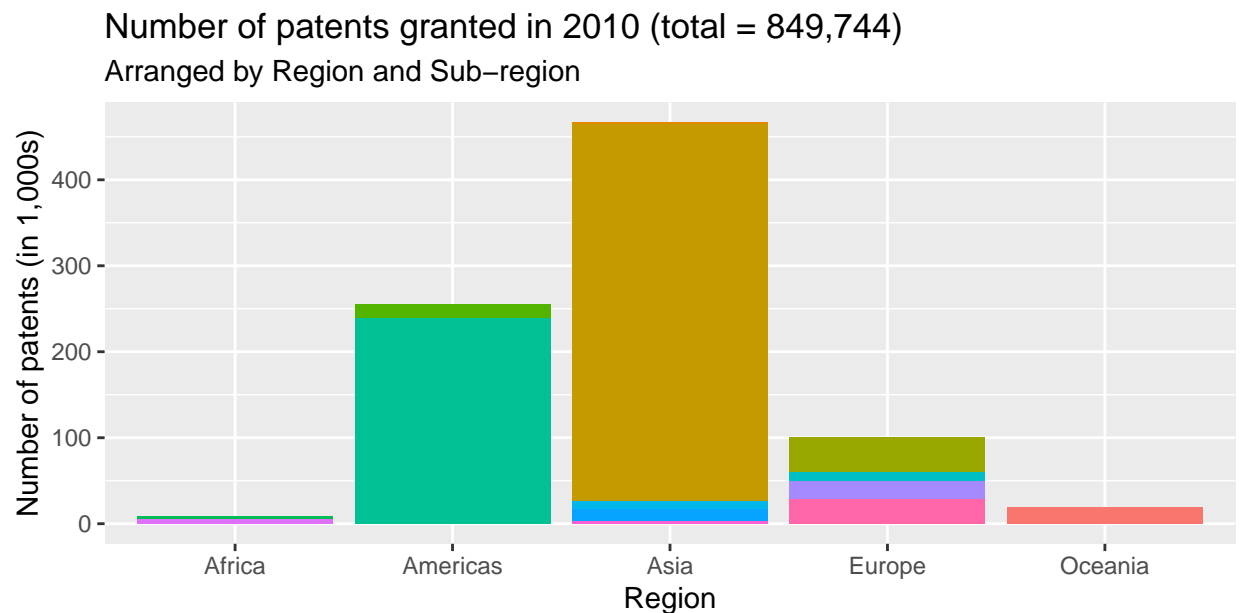
## Analyze and Visualize the Data

The four new dataframes can be displayed graphically and analyzed in order to answer the research questions.

### Patents granted in 2010

The following code chunk presents a graphical display of the number of patents granted in 2010 by region. By inspection, this graphic answers both Q1 and Q2 posed in the Introduction. Asia is the region with the largest number of patents granted, and the legend indicates the Eastern Asia sub-region accounted for the greatest number of patents granted in Asia in 2010

```
## number of patents is displayed in 1,000s to avoid scientific notation
p01 <- ggplot(un_patents_granted_2010, aes(x = region, y = number_patents/1000, fill=sub_region)) + geom_bar()
## adding labels to the plot
p01 + labs(title = "Number of patents granted in 2010 (total = 849,744)", subtitle = "Arranged by Region and Sub-region")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>



### Q1: Asia had the largest share of patents granted in 2010

The total number of patents granted to countries in Asia in 2010 was 466,526, approximately 54.9% of all patents granted in 2010. The following code chunk calculates the total number of patents granted to countries in Asia in 2010

```
## total number of patents granted in Asia in 2010
asia_total_granted_2010 <- sum(subset(un_patents_granted_2010, region == "Asia")$number_patents)
asia_total_granted_2010
```

```
## [1] 466526
```

The following code chunk calculates the percentage of patents granted to countries in Asia in 2010 of the total granted in 2010.

```
## percentage of Asian patents in 2010
asia_pct_granted_2010 <- (asia_total_granted_2010 / total_granted_2010) *100
asia_pct_granted_2010 <- round(asia_pct_granted_2010, 3)
asia_pct_granted_2010
```

```
## [1] 54.902
```

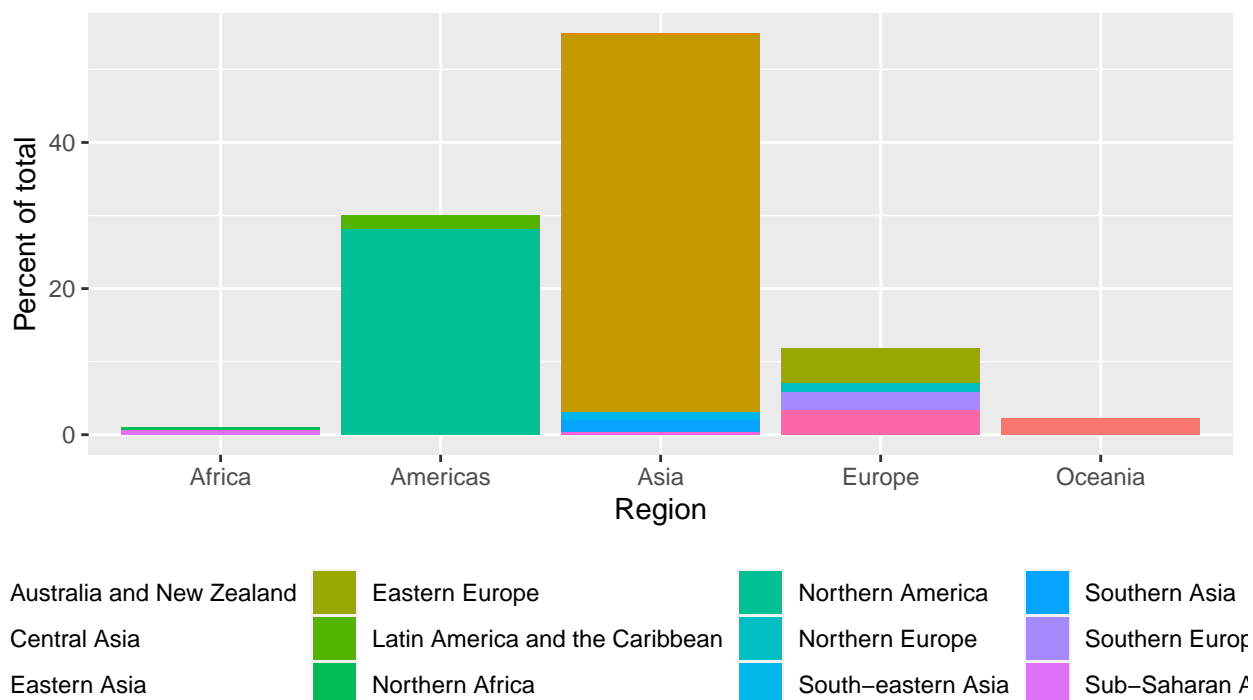
### Percentage of patents granted in Asia in 2010

The following code chunk presents a graphical display of the number of patents granted in 2010 by region as a percentage of the total number granted in 2010.

```
## percent of total patents granted in 2010
p02 <- ggplot(un_patents_granted_2010, aes(x = region, y = percent_granted, fill=sub_region)) + geom_col()
## adding labels to the plot
p02 + labs(title = "Percentage of patents granted in 2010 (total = 849,744)", subtitle = "Arranged by Region")
```

## Percentage of patents granted in 2010 (total = 849,744)

Arranged by Region and Sub-region



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

### Q2: Eastern Asia had the largest share of patents granted in Asia in 2010

The graphic indicates countries in the sub-region Eastern Asia dominated the patent activity in Asia with 438,541 patents granted in 2010, 94% of all patents granted in Asia in 2010.

Calculate the total number of patents granted in 2010 to countries in Eastern Asia:

```
## total number of patents granted in Eastern Asia in 2010
total_granted_2010_eastern_asia <- sum(subset(un_patents_granted_2010, sub_region == "Eastern Asia")$num
total_granted_2010_eastern_asia
```

```
## [1] 438541
```

Calculate the total number of patents granted in 2010 to countries in Eastern Asia as a percentage of the total number granted in Asia in 2010:

```
## percentage of Eastern Asian patents in 2010
pct_granted_2010_eastern_asia <- (total_granted_2010_eastern_asia / asia_total_granted_2010) *100
pct_granted_2010_eastern_asia <- round(pct_granted_2010_eastern_asia, 3)
pct_granted_2010_eastern_asia
```

```
## [1] 94.001
```

### Number of patents granted in Eastern Asia in 2010

```
un_patents_granted_2010_eastern_asia <- subset(un_patents_granted_2010, sub_region == "Eastern Asia",
select=country_code:sub_region_code)
```

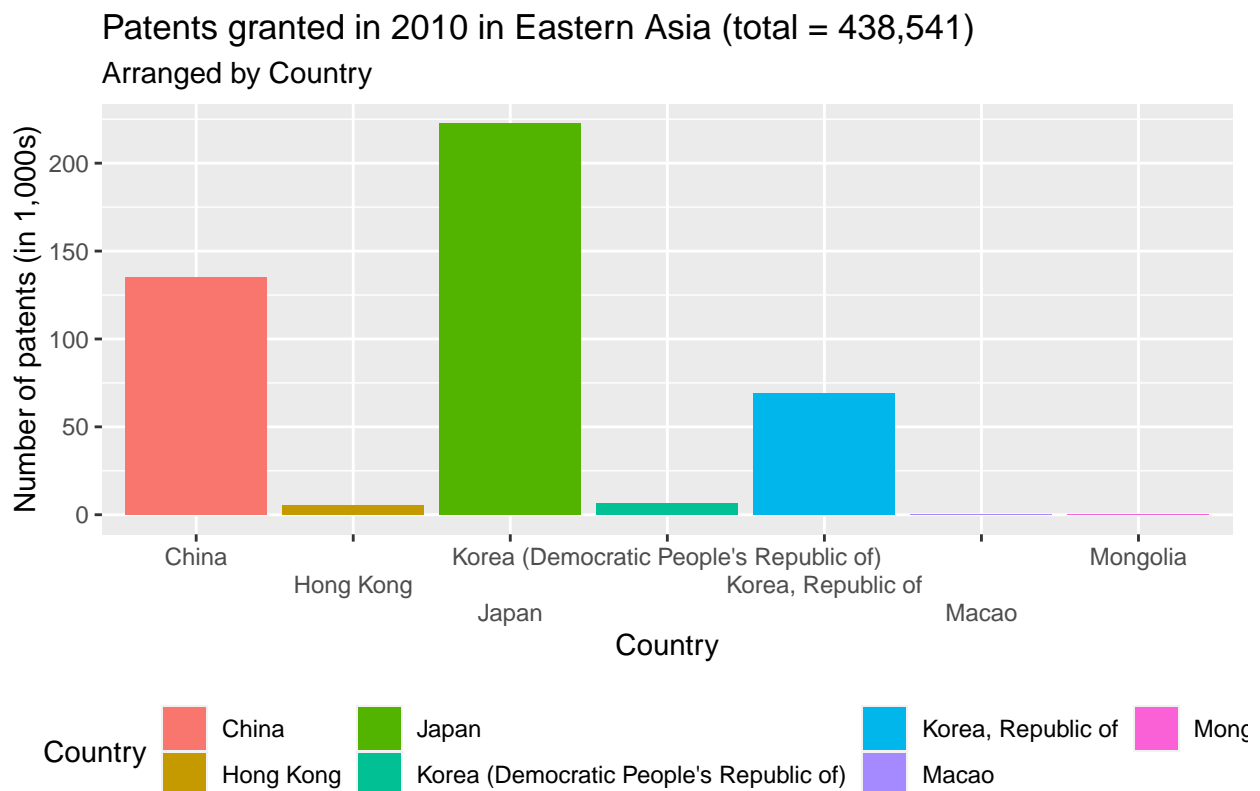
```
tibble(un_patents_granted_2010_eastern_asia)
```

```
## # A tibble: 7 x 10
##   country_~1 country iso_a~2   year number~3 series region sub_r~4 regio~5 sub_r~6
##         <dbl> <chr>   <chr>   <dbl>   <dbl> <chr>   <chr>   <chr>         <dbl>   <dbl>
## 1         156 China    CHN     2010   135110 Grant~ Asia   Easter~    142     30
## 2         344 Hong K~ HKG     2010    5353 Grant~ Asia   Easter~    142     30
## 3         392 Japan   JPN     2010  222693 Grant~ Asia   Easter~    142     30
## 4         408 Korea ~ PRK     2010    6290 Grant~ Asia   Easter~    142     30
## 5         410 Korea,~ KOR     2010   68843 Grant~ Asia   Easter~    142     30
## 6         446 Macao   MAC     2010     156 Grant~ Asia   Easter~    142     30
## 7         496 Mongol~ MNG     2010      96 Grant~ Asia   Easter~    142     30
## # ... with abbreviated variable names 1: country_code, 2: iso_alpha_3,
## #   3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code

total_granted_2010_eastern_asia <- sum(un_patents_granted_2010_eastern_asia$number_patents)
```

The following code chunk presents a graphical display of the number of patents granted to countries in Eastern Asia in 2010.

```
## number of patents is displayed in 1,000s to avoid scientific notation
## the guides function staggers the country names on the x-axis so they do not overlap
p03 <- ggplot(un_patents_granted_2010_eastern_asia, aes(x = country, y = number_patents/1000, fill=country))
## adding labels to the plot
p03 + labs(title = "Patents granted in 2010 in Eastern Asia (total = 438,541)", subtitle = "Arranged by Country")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

## Percentage of patents granted in Eastern Asia in 2010

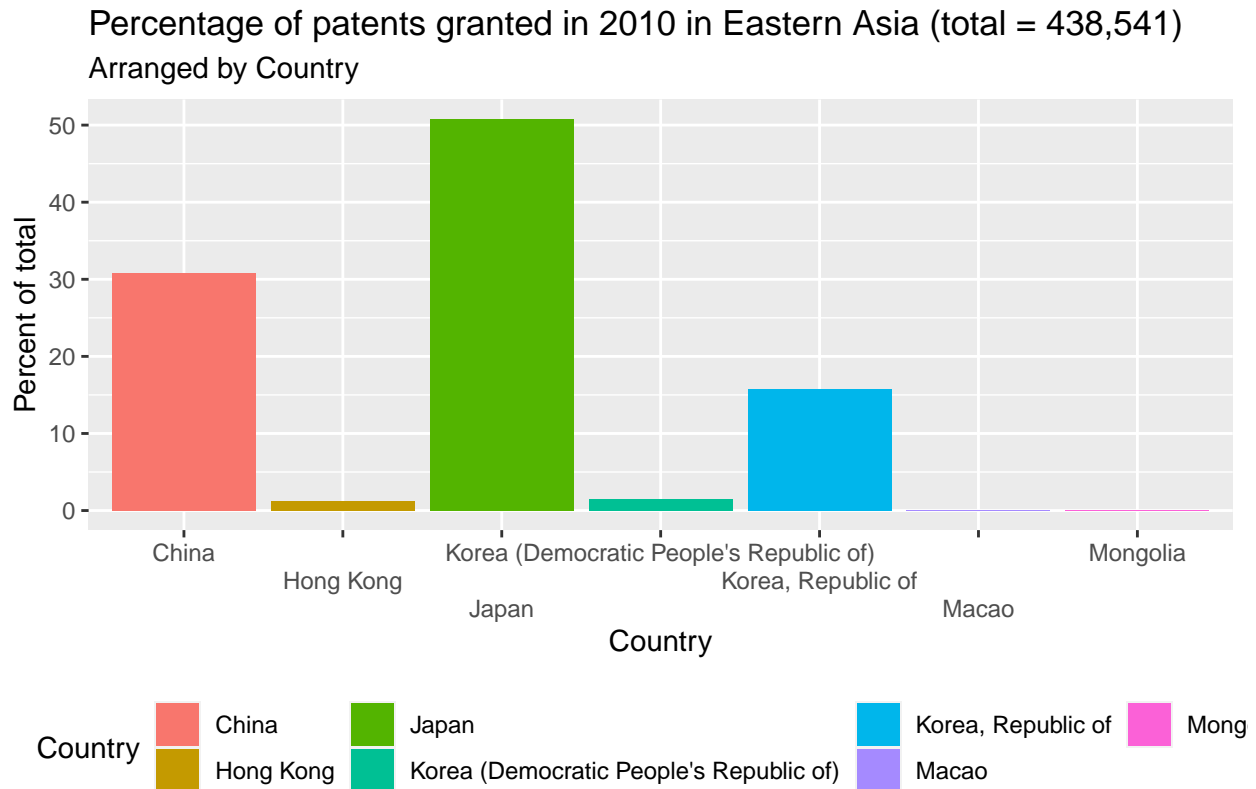
The following code chunk calculates the percentages for each country in Eastern Asia in 2010.

```
## Percent of patents granted in 2010
un_patents_granted_2010_eastern_asia$percent_granted <- (un_patents_granted_2010_eastern_asia$number_patents_granted / un_patents_granted_2010_eastern_asia$number_patents)
tibble(un_patents_granted_2010_eastern_asia)

## # A tibble: 7 x 11
##   country_~1 country iso_a~2 year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1 156 China CHN 2010 135110 Grant~ Asia Easter~ 142 30
## 2 344 Hong K~ HKG 2010 5353 Grant~ Asia Easter~ 142 30
## 3 392 Japan JPN 2010 222693 Grant~ Asia Easter~ 142 30
## 4 408 Korea ~ PRK 2010 6290 Grant~ Asia Easter~ 142 30
## 5 410 Korea,~ KOR 2010 68843 Grant~ Asia Easter~ 142 30
## 6 446 Macao MAC 2010 156 Grant~ Asia Easter~ 142 30
## 7 496 Mongol~ MNG 2010 96 Grant~ Asia Easter~ 142 30
## # ... with 1 more variable: percent_granted <dbl>, and abbreviated variable
## # names 1: country_code, 2: iso_alpha_3, 3: number_patents, 4: sub_region,
## # 5: region_code, 6: sub_region_code
```

The following code chunk presents a graphical display of the number of patents granted to countries in Eastern Asia as a percentage of the total number granted in Asia.

```
## number of patents is displayed in 1,000s to avoid scientific notation
p04 <- ggplot(un_patents_granted_2010_eastern_asia, aes(x = country, y = percent_granted, fill=country))
## adding labels to the plot
p04 + labs(title = "Percentage of patents granted in 2010 in Eastern Asia (total = 438,541)", subtitle = "Percentage of patents granted in 2010 in Eastern Asia (total = 438,541)")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

### Q3: Japan had the largest number of patents granted in Eastern Asia in 2010

A closer examination of the sub-region of Eastern Asia is required in order to answer Q3. Total number of patents granted to Japan in 2010 was 222,693, approximately 50.78% of all patents granted to countries in Eastern Asia in 2010.

Calculate the number of patents granted to Japan in 2010:

```
japan_granted_2010 <- sum(subset(un_patents_granted_2010_eastern_asia, country == "Japan")$number_patents)
japan_granted_2010
```

```
## [1] 222693
```

Calculate the number of patents granted to Japan in 2010 as a percentage of the patents granted to countries in Eastern Asia in 2010:

```
japan_pct_granted_2010 <- (japan_granted_2010 / total_granted_2010_eastern_asia) * 100
japan_pct_granted_2010 <- round(japan_pct_granted_2010, 3)
japan_pct_granted_2010
```

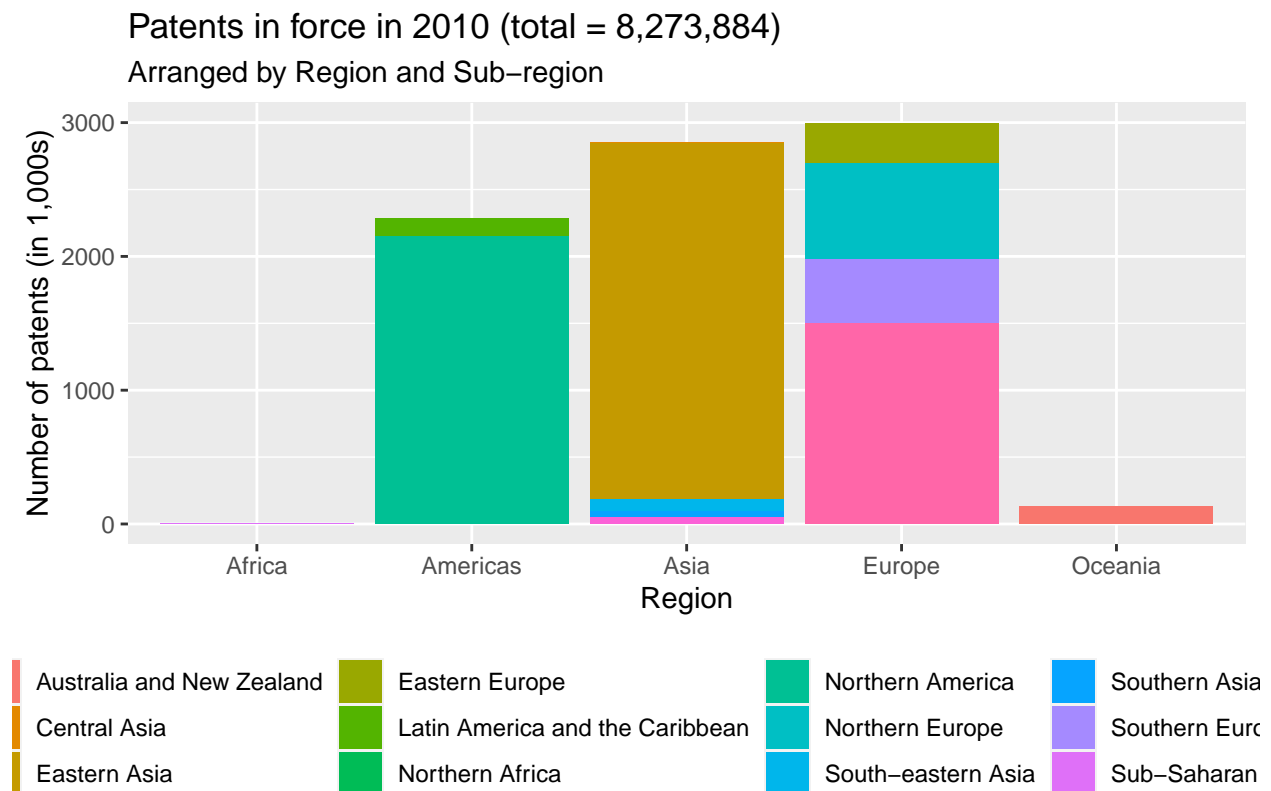
```
## [1] 50.78
```

### Patents in force in 2010

The following code chunk presents a graphical display of patents in force in 2010 by region. The graphic provides answers to Q4 and Q5 posed in the Introduction. The countries of Europe had the largest number of patents in force in 2010, and the legend indicates the sub-region Western Europe accounted for most of

those patents in force in 2010. (Note: While Western Europe was dominant in Europe, the graphic clearly indicates both sub-regions Northern America and East Asia had more patents in force than did Western Europe in 2010.)

```
## number of patents is displayed in 1,000s to avoid scientific notation
p05 <- ggplot(un_patents_in_force_2010, aes(x = region, y = number_patents/1000, fill=sub_region)) + geom_bar()
## adding labels to the plot
p05 + labs(title = "Patents in force in 2010 (total = 8,273,884)", subtitle = "Arranged by Region and Sub-region")
```

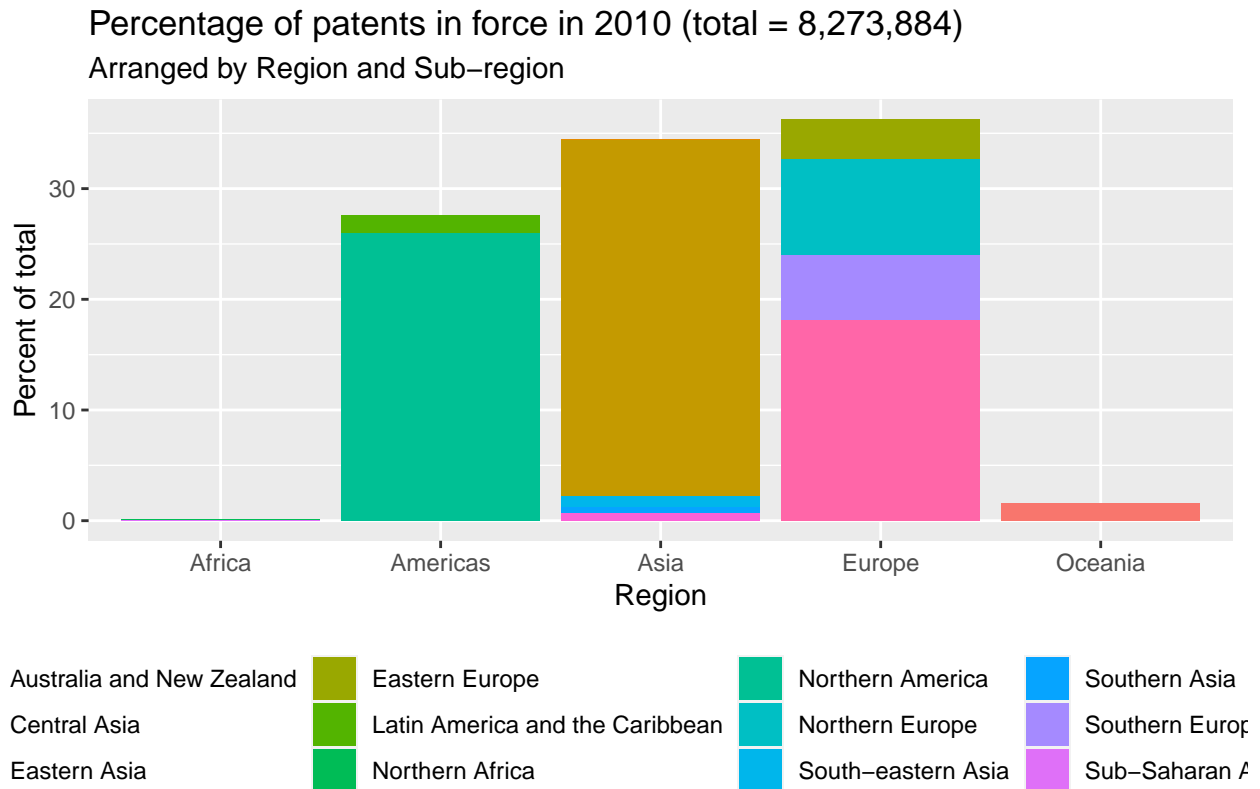


Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

## Percentage of patents in force in 2010

The following code chunk presents a graphic display of patents in force in 2010 by region as a percentage of the total number in force.

```
## percentage of patents in force in 2010
p06 <- ggplot(un_patents_in_force_2010, aes(x = region, y = percent_in_force, fill=sub_region)) + geom_bar()
## adding labels to the plot
p06 + labs(title = "Percentage of patents in force in 2010 (total = 8,273,884)", subtitle = "Arranged by Region and Sub-region")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

#### Q4: Europe had the largest share of patents in force in 2010

Europe dominated the total number of patents in force in 2010, with a total of 2,998,246 patents, approximately 36.24% of all patents in force in 2010.

Calculate the total number of patents in force in 2010 for the countries in Europe:

```
europa_total_in_force_2010 <- sum(subset(un_patents_in_force_2010, region == "Europe")$number_patents)
europa_total_in_force_2010
```

```
## [1] 2998246
```

Calculate the total number of patents in force in 2010 for the countries in Europe as a percentage of the total number in force in 2010.

```
europa_pct_in_force_2010 <- (europa_total_in_force_2010 / total_in_force_2010) * 100
europa_pct_in_force_2010 <- round(europa_pct_in_force_2010, 3)
europa_pct_in_force_2010
```

```
## [1] 36.237
```

#### Q5: Western Europe had the largest share of patents in force in Europe in 2010

Western Europe was the dominant sub-region in Europe for patents in force in 2010, with 1,496,897 patents, 49.93% of all patents in force in Europe in 2010.

Calculate the total number of patents in force in 2010 belonging to the countries of Western Europe:

```
## total number of patents in force in 2010 in Western Europe
total_in_force_2010_western_europe <- sum(subset(un_patents_in_force_2010, sub_region == "Western Europe"))
total_in_force_2010_western_europe
```

```
## [1] 1496897
```

Calculate the total number of patents in force in 2010 in Western Europe as a percentage of the total number granted in Europe in 2010:

```
## percentage of Western Europe patents in force in 2010
pct_in_force_2010_western_europe <- (total_in_force_2010_western_europe / europe_total_in_force_2010) * 100
pct_in_force_2010_western_europe <- round(pct_in_force_2010_western_europe, 3)
pct_in_force_2010_western_europe
```

```
## [1] 49.926
```

A closer examination of the sub-region of Western Europe is required in order to answer Q6. The following code chunk creates a subset of data for Western Europe in order to identify the country with the largest share of patents in force in 2010.

```
un_patents_in_force_2010_western_europe <- subset(un_patents_in_force_2010, sub_region == "Western Europe")
select=country_code:sub_region_code)
tibble(un_patents_in_force_2010_western_europe)
```

```
## # A tibble: 8 x 10
##   country_~1 country iso_a~2   year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      40 Austria AUT    2010  102113 Paten~ Europe Wester~    150    155
## 2      56 Belgium BEL    2010   89999 Paten~ Europe Wester~    150    155
## 3     250 France  FRA    2010  435915 Paten~ Europe Wester~    150    155
## 4     276 Germany DEU    2010  514046 Paten~ Europe Wester~    150    155
## 5     442 Luxemb~ LUX    2010   42805 Paten~ Europe Wester~    150    155
## 6     492 Monaco MCO    2010   53859 Paten~ Europe Wester~    150    155
## 7     528 Nether~ NLD    2010  135127 Paten~ Europe Wester~    150    155
## 8     756 Switze~ CHE    2010  123033 Paten~ Europe Wester~    150    155
## # ... with abbreviated variable names 1: country_code, 2: iso_alpha_3,
## # 3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
```

## Number of patents in force in Western Europe in 2010

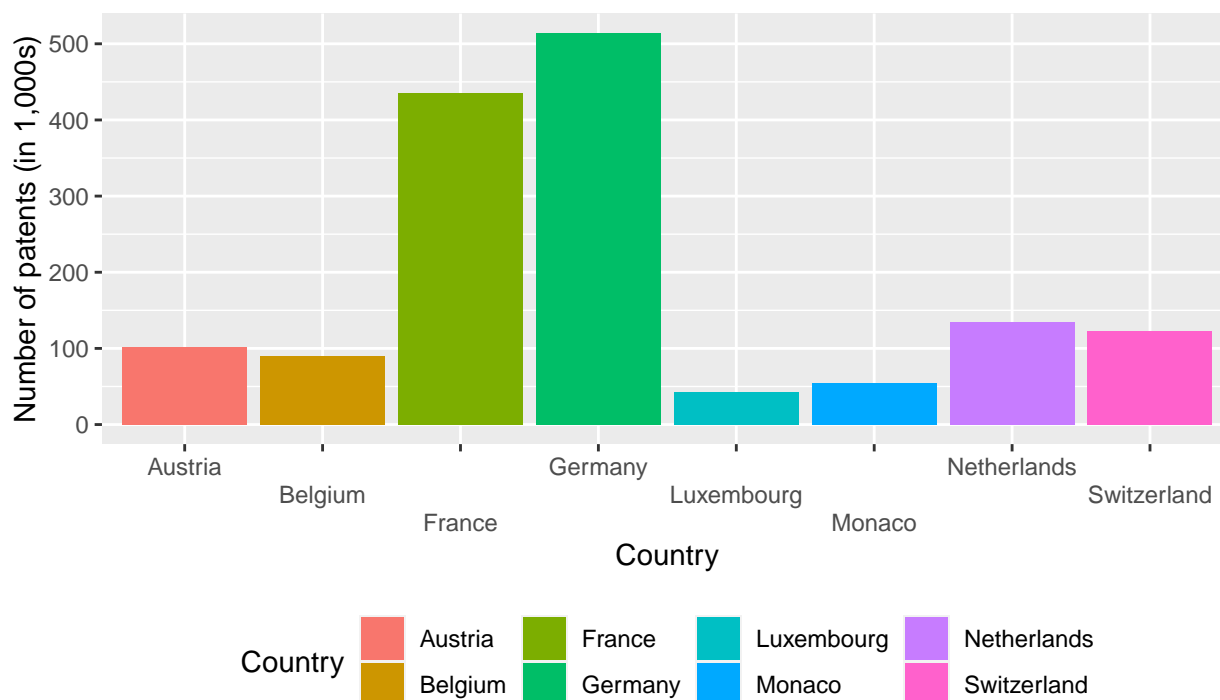
The following code chunk presents a graphical display of the patents in force in 2010 in the sub-region Western Europe by country. The legend clearly indicates that Germany had the largest number of patents in force in 2010 of all countries in Western Europe.

```
## number of patents is displayed in 1,000s to avoid scientific notation
## the guides function staggers the country names on the x-axis so they do not overlap
p07 <- ggplot(un_patents_in_force_2010_western_europe, aes(x = country, y = number_patents/1000, fill=country))
## adding labels to the plot
p07 + labs(title = "Patents in force in Western Europe in 2010 (total = 2,998,246)", subtitle = "Arranged by country")
```



## Patents in force in Western Europe in 2010 (total = 2,998,246)

Arranged by Country



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

## Percentage of patents in force in Western Europe in 2010

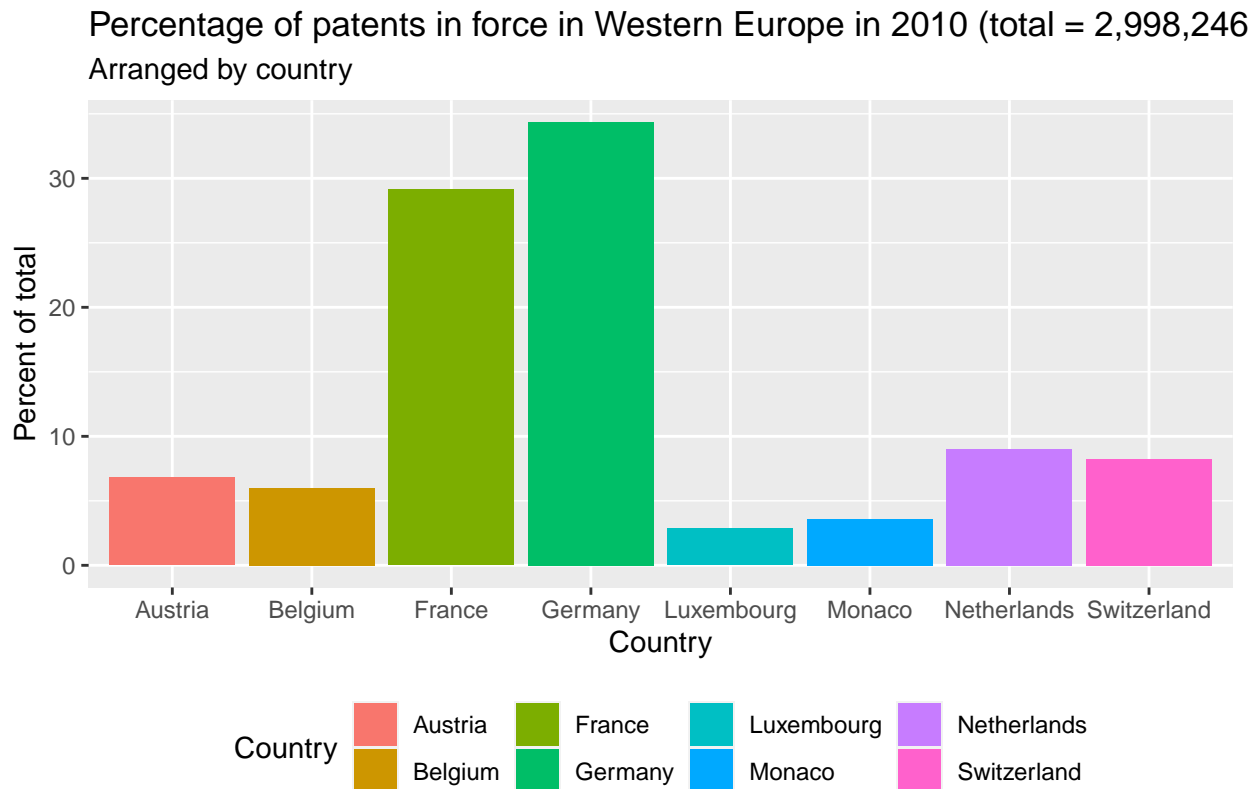
The following code chunk calculates the number of patents in force in 2010 for each country in Western Europe as a percentage of all patents in force in Western Europe in 2010.

```
## Percent of patents in force 2010
un_patents_in_force_2010_western_europe$percent_in_force <- (un_patents_in_force_2010_western_europe$num
un_patents_in_force_2010_western_europe$percent_in_force <- round(un_patents_in_force_2010_western_europe
tibble(un_patents_in_force_2010_western_europe)
```

```
## # A tibble: 8 x 11
##   country_~1 country iso_a~2 year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      40 Austria AUT      2010 102113 Paten~ Europe Wester~      150      155
## 2      56 Belgium BEL      2010 899999 Paten~ Europe Wester~      150      155
## 3     250 France FRA      2010 435915 Paten~ Europe Wester~      150      155
## 4     276 Germany DEU      2010 514046 Paten~ Europe Wester~      150      155
## 5     442 Luxemb~ LUX      2010 42805 Paten~ Europe Wester~      150      155
## 6     492 Monaco MCO      2010 53859 Paten~ Europe Wester~      150      155
## 7     528 Nether~ NLD      2010 135127 Paten~ Europe Wester~      150      155
## 8     756 Switze~ CHE      2010 123033 Paten~ Europe Wester~      150      155
## # ... with 1 more variable: percent_in_force <dbl>, and abbreviated variable
## # names 1: country_code, 2: iso_alpha_3, 3: number_patents, 4: sub_region,
## # 5: region_code, 6: sub_region_code
```

The following code chunk presents a graphical display of the patents in force in 2010 for each country in Western Europe as a percentage of the total number of patents in force in Western Europe in 2010.

```
## percentage of patents in force in Western Europe in 2010
p08 <- ggplot(un_patents_in_force_2010_western_europe, aes(x = country, y = percent_in_force, fill=country))
## adding labels to the plot
p08 + labs(title = "Percentage of patents in force in Western Europe in 2010 (total = 2,998,246)", subtitle = "Arranged by country")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

#### Q6: Germany had the largest share of patents in force in Western Europe in 2010

Germany dominated the total number of patents in force in Western Europe in 2010 with 514,046, approximately 34.34% of all patents in force in Western Europe in 2010

Calculate the number of patents Germany had in force in 2010:

```
## Number of German patents in force in 2010
germany_in_force_2010 <- sum(subset(un_patents_in_force_2010_western_europe, country == "Germany")$number_of_patents_in_force_2010)
```

```
## [1] 514046
```

Calculate the number of patents Germany had in force in 2010 as a percentage of all patents in force in Western Europe in 2010:

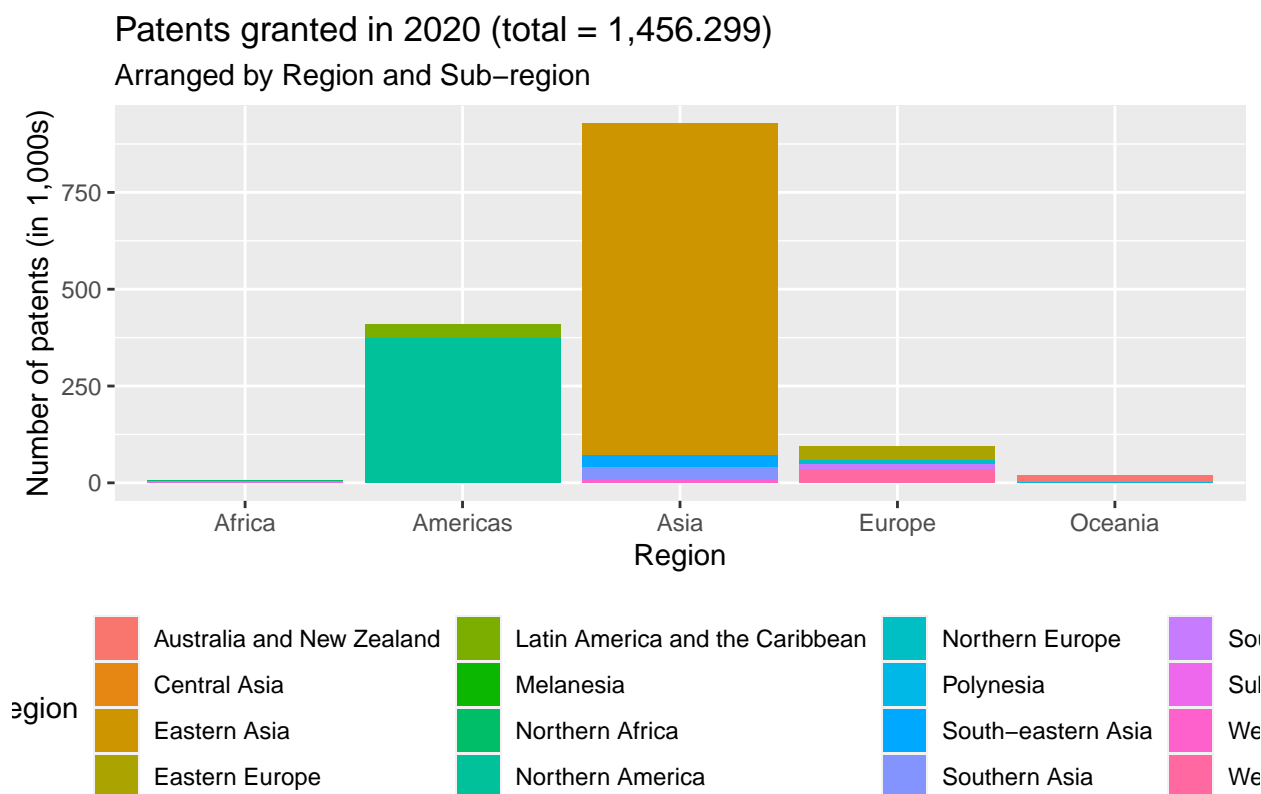
```
## German patents in force in 2010 as a percentage of Western European patents in force
germany_pct_in_force_2010 <- (germany_in_force_2010 / total_in_force_2010_western_europe) * 100
germany_pct_in_force_2010 <- round(germany_pct_in_force_2010, 3)
germany_pct_in_force_2010
```

```
## [1] 34.341
```

## Patents granted in 2020

The following code chunk presents a graphical display of patents granted in 2020 by region. This graphic answers both Q7 and Q8 posed in the introduction. Asia is the region with the largest number of patents granted, and the legend indicates the Eastern Asia sub-region accounted for the greatest number of patents granted in 2020.

```
## number of patents is displayed in 1,000s to avoid scientific notation
p09 <- ggplot(un_patents_granted_2020, aes(x = region, y = number_patents/1000, fill=sub_region)) + geom_bar()
## adding labels to the plot
p09 + labs(title = "Patents granted in 2020 (total = 1,456.299)", subtitle = "Arranged by Region and Sub-region")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

### Q7: Asia had the largest share of patents granted in 2020

Total number of patents granted to countries in Asia in 2020 was 927,612, approximately 63.7% of all patents granted in 2020.

Calculate the total number of patents granted to countries in Asia in 2020

```
## total number of patents granted in Asia in 2020
asia_total_granted_2020 <- sum(subset(un_patents_granted_2020, region == "Asia")$number_patents)
asia_total_granted_2020
```

```
## [1] 927612
```

Calculate the number of patents granted to countries in Asia in 2020 as a percentage of all patents granted in 2020

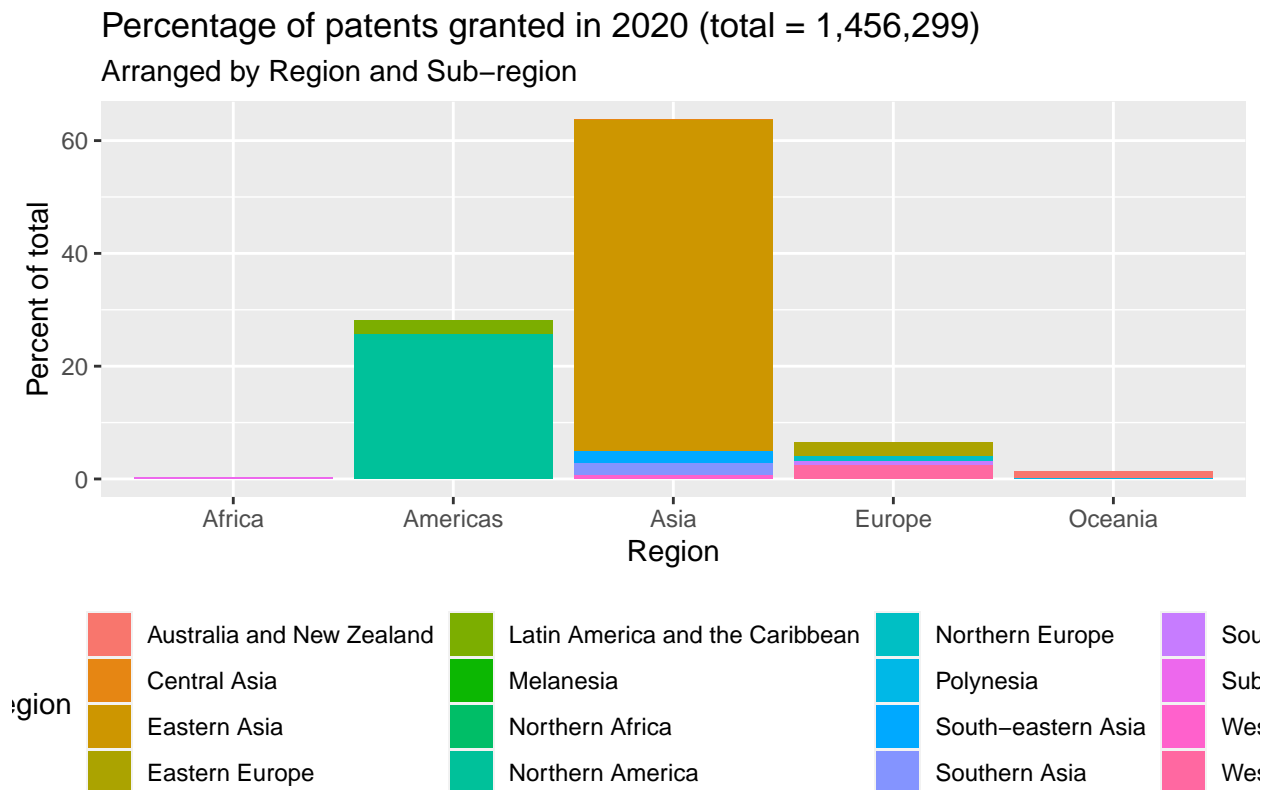
```
## percentage of patents granted in Asia in 2020
asia_pct_granted_2020 <- (asia_total_granted_2020 / total_granted_2020) * 100
asia_pct_granted_2020 <- round(asia_pct_granted_2020, 3)
asia_pct_granted_2020
```

```
## [1] 63.697
```

### Percentage of patents granted in 2020

The following code chunk presents a graphical display of the number of patents granted in 2020 by region as a percentage of the total number granted in 2020.

```
## percent of total patents granted in 2020
p10 <- ggplot(un_patents_granted_2020, aes(x = region, y = percent_granted, fill=sub_region)) + geom_col()
## adding labels to the plot
p10 + labs(title = "Percentage of patents granted in 2020 (total = 1,456,299)", subtitle = "Arranged by Region and Sub-region")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

### Q8: Eastern Asia had the largest share of patents granted in Asia in 2020

Eastern Asia was the dominant sub-region in Asia for patents granted in 2020: 856,303 patents, approximately 92.31% of the total granted in Asia in 2020.

Calculate the total number of patents granted in 2020 to countries in Eastern Asia:

```
## total number of patents granted in Eastern Asia in 2020
total_granted_2020_eastern_asia <- sum(subset(un_patents_granted_2020, sub_region == "Eastern Asia")$num)
total_granted_2020_eastern_asia
```

```
## [1] 856303
```

Calculate the total number of patents granted in 2020 to countries in Eastern Asia as a percentage of the total number granted in Asia in 2020:

```
## percentage of Eastern Asian patents in 2010
```

```
pct_granted_2020_eastern_asia <- (total_granted_2020_eastern_asia / asia_total_granted_2020) *100
pct_granted_2020_eastern_asia <- round(pct_granted_2020_eastern_asia, 3)
pct_granted_2020_eastern_asia
```

```
## [1] 92.313
```

A closer examination of the sub-region of Eastern Asia is required in order to answer Q9.

```
un_patents_granted_2020_eastern_asia <- subset(un_patents_granted_2020, sub_region == "Eastern Asia",
select=country_code:sub_region_code)
tibble(un_patents_granted_2020_eastern_asia)
```

```
## # A tibble: 7 x 10
```

```
##   country_~1 country iso_a~2   year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1      156 China  CHN    2020  530127 Grant~ Asia  Easter~    142    30
## 2      344 Hong K~ HKG    2020    7658 Grant~ Asia  Easter~    142    30
## 3      392 Japan  JPN    2020  179383 Grant~ Asia  Easter~    142    30
## 4      408 Korea ~ PRK    2020    4227 Grant~ Asia  Easter~    142    30
## 5      410 Korea,~ KOR    2020  134766 Grant~ Asia  Easter~    142    30
## 6      446 Macao  MAC    2020     18 Grant~ Asia  Easter~    142    30
## 7      496 Mongol~ MNG    2020    124 Grant~ Asia  Easter~    142    30
```

```
## # ... with abbreviated variable names 1: country_code, 2: iso_alpha_3,
```

```
## #   3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
```

```
total_granted_2020_eastern_asia <- sum(un_patents_granted_2020_eastern_asia$number_patents)
```

## Number of patents granted in Eastern Asia in 2020

The following code chunk presents a graphical display of the number of patents granted to countries in Eastern Asia in 2020.

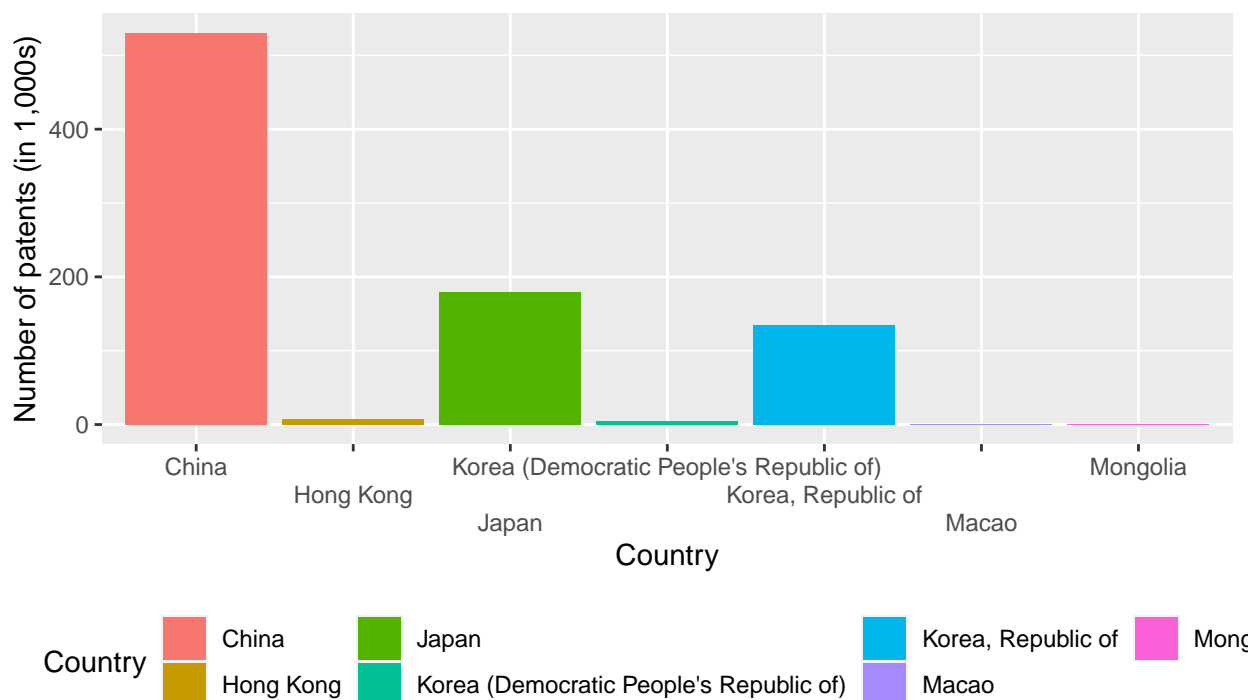
```
## number of patents is displayed in 1,000s to avoid scientific notation
```

```
p11 <- ggplot(un_patents_granted_2020_eastern_asia, aes(x = country, y = number_patents/1000, fill=country))
## adding labels to the plot
```

```
p11 + labs(title = "Patents granted in 2020 in Eastern Asia (total = 856,303)", subtitle = "Arranged by country")
```

## Patents granted in 2020 in Eastern Asia (total = 856,303)

Arranged by Country



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

## Percentages of patents granted in Eastern Asia in 2020

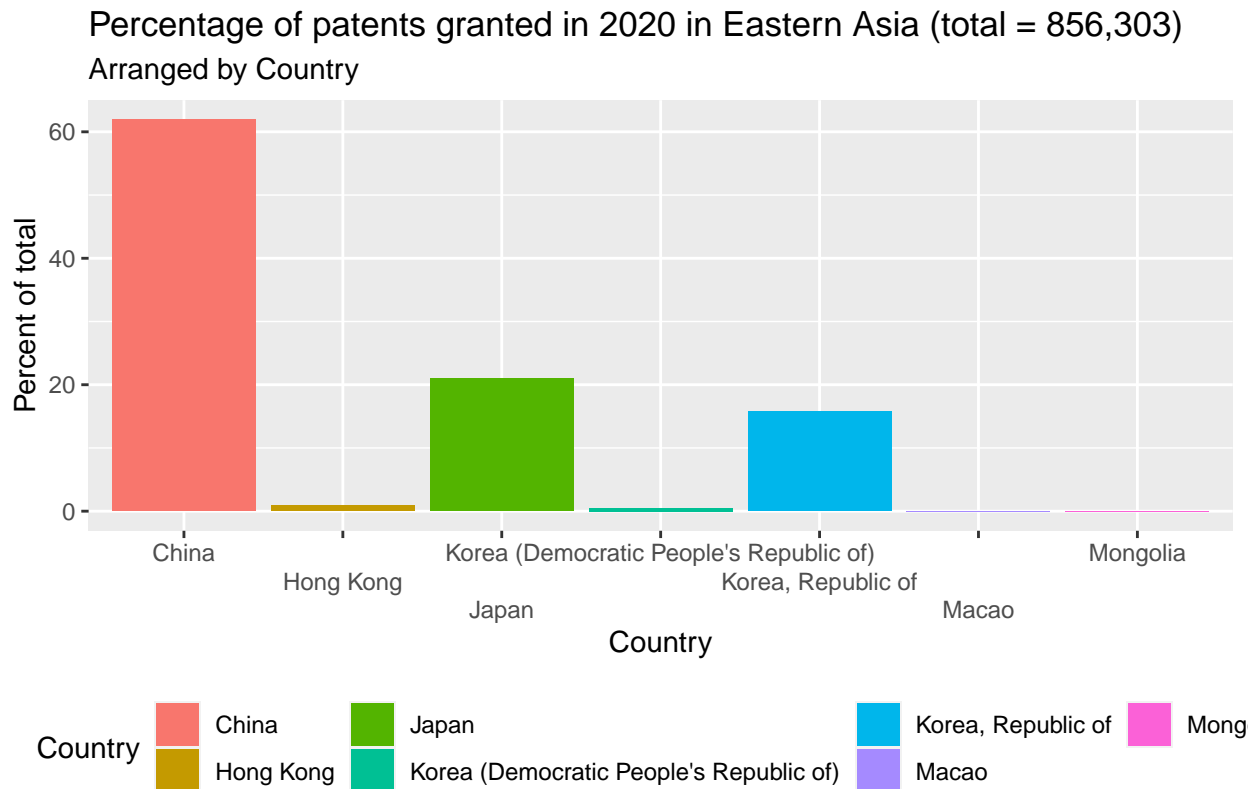
The following code chunk calculates the percentages for each country in Eastern Asia in 2020.

```
## Percent of patents granted in 2020
un_patents_granted_2020_eastern_asia$percent_granted <- (un_patents_granted_2020_eastern_asia$number_patents / 856303) * 100
un_patents_granted_2020_eastern_asia$percent_granted <- round(un_patents_granted_2020_eastern_asia$percent_granted, 2)
tibble(un_patents_granted_2020_eastern_asia)
```

```
## # A tibble: 7 x 11
##   country_code country iso_alpha3 year number_patents series region sub_region region_code sub_region_code
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1 156 China CHN 2020 530127 Grant Asia Easter Asia 142 30
## 2 344 Hong Kong HKG 2020 7658 Grant Asia Easter Asia 142 30
## 3 392 Japan JPN 2020 179383 Grant Asia Easter Asia 142 30
## 4 408 Korea DPR PRK 2020 4227 Grant Asia Easter Asia 142 30
## 5 410 Korea, Rep. KOR 2020 134766 Grant Asia Easter Asia 142 30
## 6 446 Macao MAC 2020 18 Grant Asia Easter Asia 142 30
## 7 496 Mongolia MNG 2020 124 Grant Asia Easter Asia 142 30
## # ... with 1 more variable: percent_granted <dbl>, and abbreviated variable
## # names 1: country_code, 2: iso_alpha3, 3: number_patents, 4: sub_region,
## # 5: region_code, 6: sub_region_code
```

The following code chunk presents a graphical display of the number of patents granted to countries in Eastern Asia as a percentage of the total number granted in Asia.

```
## number of patents is displayed in 1,000s to avoid scientific notation
p12 <- ggplot(un_patents_granted_2020_eastern_asia, aes(x = country, y = percent_granted, fill=country))
## adding labels to the plot
p12 + labs(title = "Percentage of patents granted in 2020 in Eastern Asia (total = 856,303)", subtitle =
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

#### Q9: China had the largest number of patents granted in Eastern Asia in 2020

Total number of patents granted to China in 2020 was 530,127, approximately 61.91% of all patents granted to countries in Eastern Asia.

Calculate the number of patents granted to China in 2020:

```
## Number of patents granted to China in 2020
china_granted_2020 <- sum(subset(un_patents_granted_2020_eastern_asia, country == "China")$number_patents)
china_granted_2020
```

```
## [1] 530127
```

Calculate the number of patents granted to China in 2020 as a percentage of the patents granted to countries in Eastern Asia in 2020:

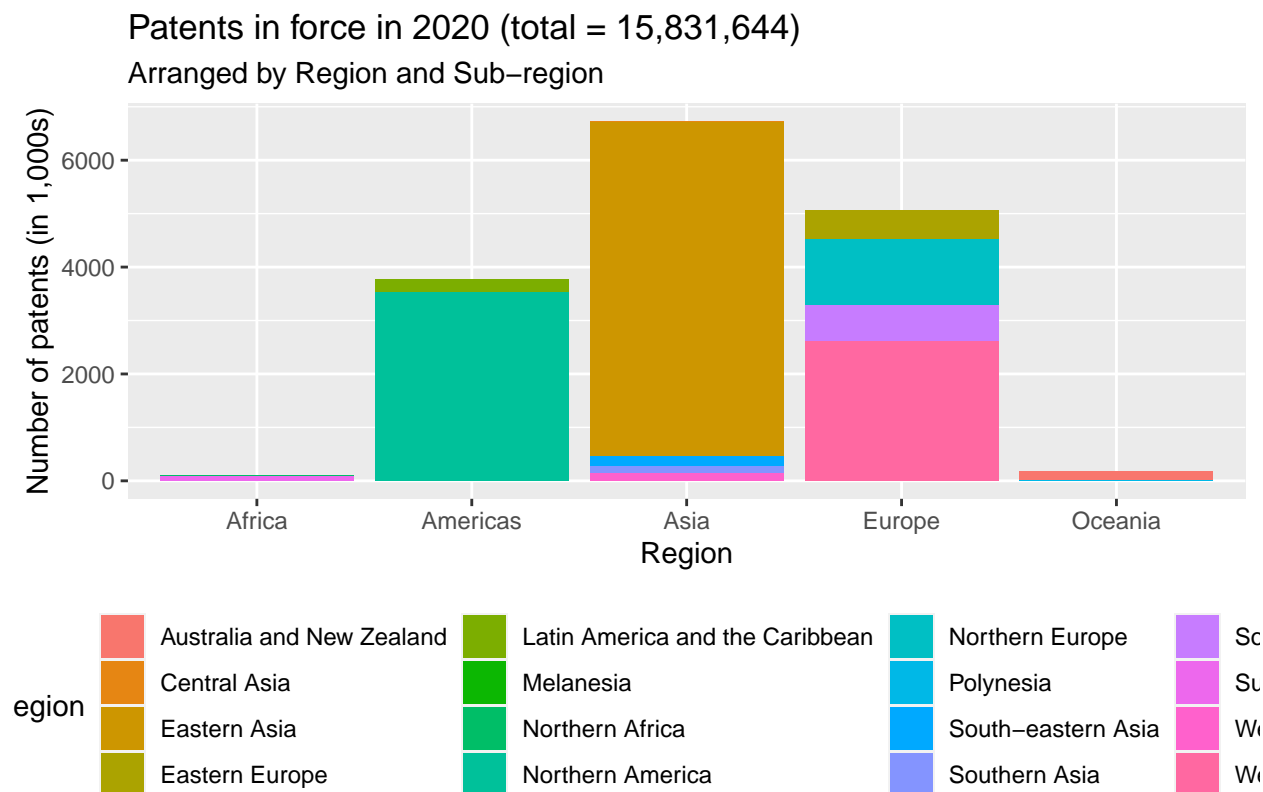
```
## Percentage of Eastern Asia patents granted to China in 2020
china_pct_granted_2020 <- (china_granted_2020 / total_granted_2020_eastern_asia) * 100
china_pct_granted_2020 <- round(china_pct_granted_2020, 3)
china_pct_granted_2020
```

```
## [1] 61.909
```

## Patents in force in 2020

The following code chunk presents a graphical display of patents in force in 2020 by region. The graphic provides answers to Q10 and Q11 posed in the introduction. The countries of Asia had the largest total number of patents in force in 2020, and the legend indicates the sub-region Eastern Asia accounted for most of those patents in force in 2020.

```
## number of patents is displayed in 1,000s to avoid scientific notation
p13 <- ggplot(un_patents_in_force_2020, aes(x = region, y = number_patents/1000, fill=sub_region)) + geom_bar()
## adding labels to the plot
p13 + labs(title = "Patents in force in 2020 (total = 15,831,644)", subtitle = "Arranged by Region and Sub-region")
```



### Q10: Asia had the largest share of patents in force in 2020

Asia dominated the total number of patents in force in 2020, with a total of 6,721,553 patents, approximately 42.46% of all patents in force in 2020.

Calculate the total number of patents in force in 2020 belonging to countries in Asia:

```
asia_in_force_2020 <- sum(subset(un_patents_in_force_2020, region == "Asia")$number_patents)
asia_in_force_2020
```

```
## [1] 6721553
```

Calculate the number of patents in force in 2020 belonging to countries in Asia as a percentage of the total number in force in 2020:



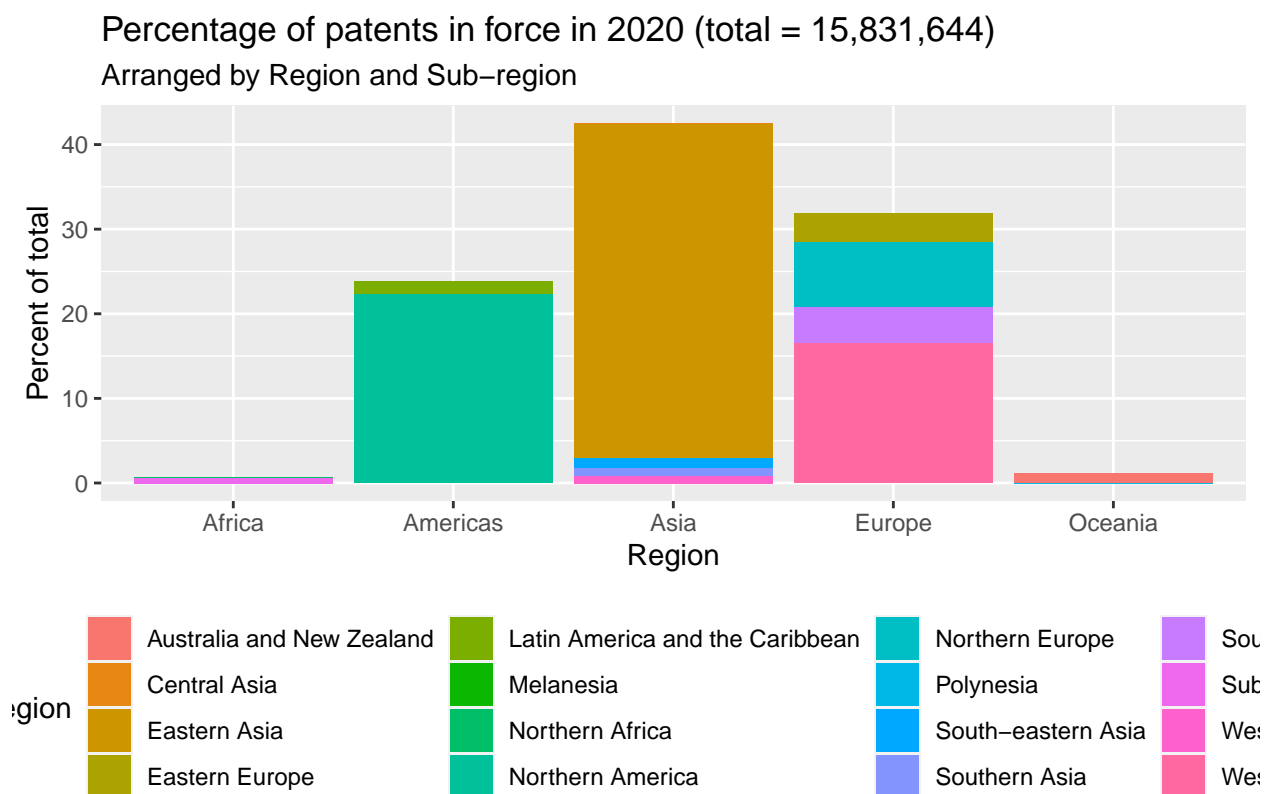
```
asia_pct_in_force_2020 <- (asia_in_force_2020 / total_in_force_2020) * 100
asia_pct_in_force_2020 <- round(asia_pct_in_force_2020, 3)
asia_pct_in_force_2020
```

```
## [1] 42.456
```

### Percentage of patents in force in 2020

The following code chunk presents a graphic display of patents in force in 2020 by region as a percentage of the total number in force.

```
## percentage of patents in force in 2020
p14 <- ggplot(un_patents_in_force_2020, aes(x = region, y = percent_in_force, fill=sub_region)) + geom_bar()
## adding labels to the plot
p14 + labs(title = "Percentage of patents in force in 2020 (total = 15,831,644)", subtitle = "Arranged by Region and Sub-region")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

### Q11: Eastern Asia had the largest share of patents in force in Asia in 2020

Eastern Asia had the largest share of patents in force in Asia in 2020: 6,249,001 patents, approximately 92.97% of all patents in force in Asia in 2020

Calculate the total number of patents in force in 2020 belonging to the countries of Eastern Asia:

```
## total number of patents in force in 2020 in Eastern Asia
total_in_force_2020_eastern_asia <- sum(subset(un_patents_in_force_2020, sub_region == "Eastern Asia")$patents_in_force_2020)
total_in_force_2020_eastern_asia
```

```
## [1] 6249001
```

Calculate the total number of patents in force in 2020 in Eastern Asia as a percentage of the total number in force in Asia in 2020:

```
## percentage of Eastern Asia patents in force in 2020
pct_in_force_2020_eastern_asia <- (total_in_force_2020_eastern_asia / asia_in_force_2020) *100
pct_in_force_2020_eastern_asia <- round(pct_in_force_2020_eastern_asia, 3)
pct_in_force_2020_eastern_asia
```

```
## [1] 92.97
```

A closer examination of the sub-region of Eastern Asia is required in order to answer Q12 posed in the introduction. The following code chunk creates a subset of data for Eastern Asia in order to identify the country with the largest share of patents in force in 2020.

```
un_patents_in_force_2020_eastern_asia <- subset(un_patents_in_force_2020, sub_region == "Eastern Asia",
select=country_code:sub_region_code)
tibble(un_patents_in_force_2020_eastern_asia)
```

```
## # A tibble: 6 x 10
##   country_~1 country iso_a~2   year numbe~3 series region sub_r~4 regio~5 sub_r~6
##   <dbl> <chr>   <chr>   <dbl>   <dbl> <chr>   <chr>   <chr>   <dbl>   <dbl>
## 1      156 China   CHN     2020 3057844 Paten~ Asia   Easter~    142     30
## 2      344 Hong K~ HKG     2020  53726 Paten~ Asia   Easter~    142     30
## 3      392 Japan   JPN     2020 2039040 Paten~ Asia   Easter~    142     30
## 4      410 Korea,~ KOR     2020 1096721 Paten~ Asia   Easter~    142     30
## 5      446 Macao   MAC     2020    353 Paten~ Asia   Easter~    142     30
## 6      496 Mongol~ MNG     2020   1317 Paten~ Asia   Easter~    142     30
## # ... with abbreviated variable names 1: country_code, 2: iso_alpha_3,
## #   3: number_patents, 4: sub_region, 5: region_code, 6: sub_region_code
total_in_force_2020_eastern_asia <- sum(un_patents_in_force_2020_eastern_asia$number_patents)
```

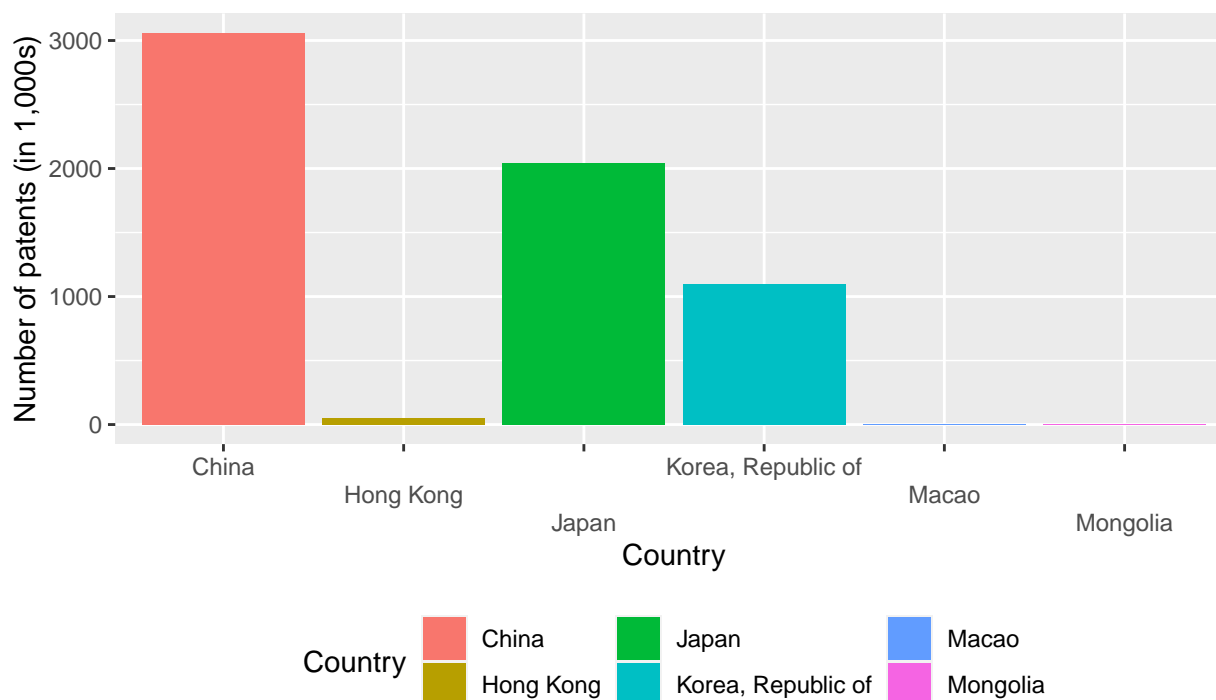
## Number of patents in force in Eastern Asia in 2020

The following code chunk provides a graphical display of the number of patents in force in Eastern Asia in 2020

```
## number of patents is displayed in 1,000s to avoid scientific notation
p15 <- ggplot(un_patents_in_force_2020_eastern_asia, aes(x = country, y = number_patents/1000, fill=country_code))
## adding labels to the plot
p15 + labs(title = "Patents in force in 2020 in Eastern Asia (total = 6,249,001)", subtitle = "Arranged by country")
```

## Patents in force in 2020 in Eastern Asia (total = 6,249,001)

Arranged by Country



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

## Percentages of patents in force in Eastern Asia in 2020

The following code chunk calculates for each country in the sub-region Eastern Asia the percentage of the total patents in force in Eastern Asia in 2020 and adds the column to the dataframe.

```
un_patents_in_force_2020_eastern_asia$percent_in_force <- (un_patents_in_force_2020_eastern_asia$number_
un_patents_in_force_2020_eastern_asia$percent_in_force <- round(un_patents_in_force_2020_eastern_asia$number_
tibble(un_patents_in_force_2020_eastern_asia)
```

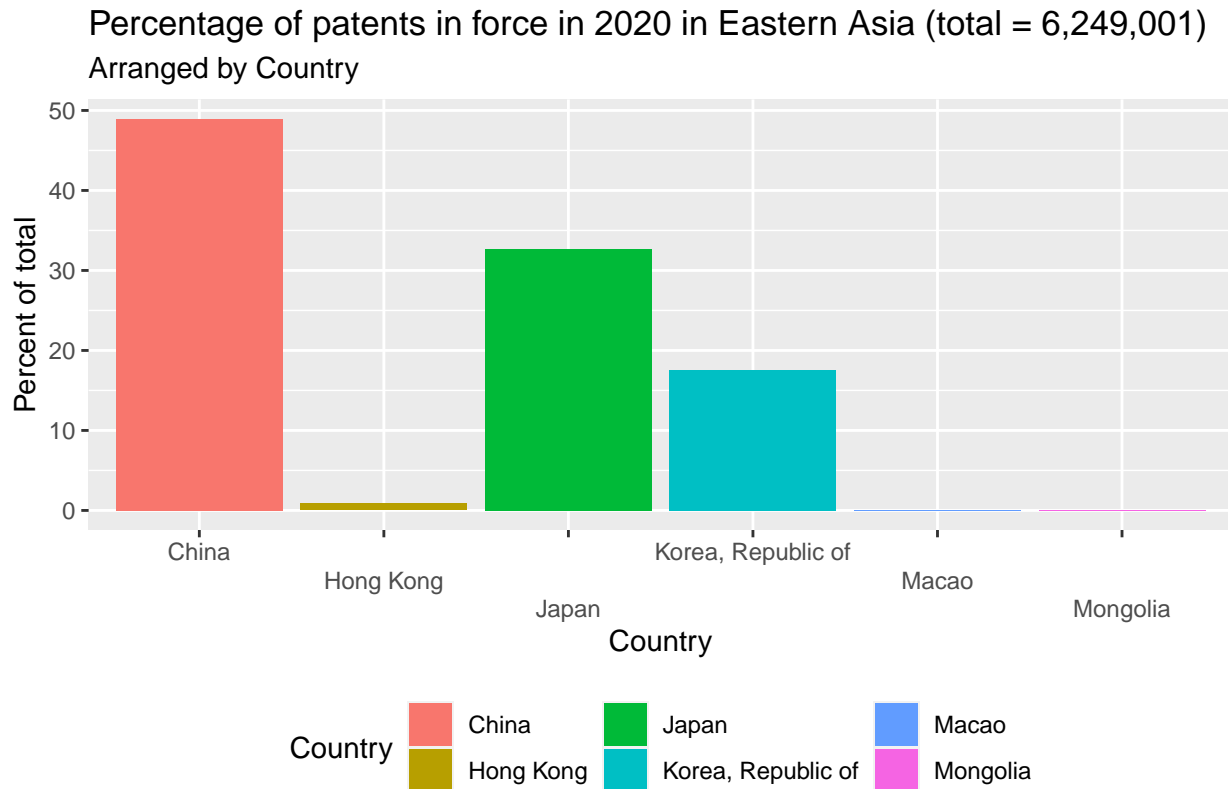
```
## # A tibble: 6 x 11
##   country_1 country iso_a2 year number series region sub_r4 regio5 sub_r6
##   <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl>
## 1 156 China CHN 2020 3057844 Paten~ Asia Easter~ 142 30
## 2 344 Hong K~ HKG 2020 53726 Paten~ Asia Easter~ 142 30
## 3 392 Japan JPN 2020 2039040 Paten~ Asia Easter~ 142 30
## 4 410 Korea,~ KOR 2020 1096721 Paten~ Asia Easter~ 142 30
## 5 446 Macao MAC 2020 353 Paten~ Asia Easter~ 142 30
## 6 496 Mongol~ MNG 2020 1317 Paten~ Asia Easter~ 142 30
## # ... with 1 more variable: percent_in_force <dbl>, and abbreviated variable
## # names 1: country_code, 2: iso_alpha_3, 3: number_patents, 4: sub_region,
## # 5: region_code, 6: sub_region_code
```

The following code chunk presents a graphical display of the percentage of patents in force in 2020 for each country in the sub-region Eastern Asia.

```
## number of patents is displayed in 1,000s to avoid scientific notation
p16 <- ggplot(un_patents_in_force_2020_eastern_asia, aes(x = country, y = percent_in_force, fill=country_
```

```
## adding labels to the plot
```

```
p16 + labs(title = "Percentage of patents in force in 2020 in Eastern Asia (total = 6,249,001)", subtitle = "Arranged by Country")
```



Source: UNdata, Science and technology, Patents (updated 20221018), <http://data.un.org>

## Q12: China had the largest share of patents in force in Eastern Asia in 2020

China dominated the patents in force in Eastern Asia in 2020: 3,057,844 patents, approximately 48.93% of all patents in force in Eastern Asia in 2020.

Calculate the number of patents in force in 2020 for China:

```
china_in_force_2020 <- sum(subset(un_patents_in_force_2020_eastern_asia, country == "China")$number_patents)
china_in_force_2020
```

```
## [1] 3057844
```

Calculate the number of patents in force in 2020 for China as a percentage of the number of patents in force in Eastern Asia in 2020:

```
china_pct_in_force_2020 <- (china_in_force_2020 / total_in_force_2020_eastern_asia) * 100
china_pct_in_force_2020 <- round(china_pct_in_force_2020, 3)
china_pct_in_force_2020
```

```
## [1] 48.933
```

## Conclusion

The analysis has identified the following significant trends:

- Asia was dominant in terms of patents granted both in 2010 and 2020
- Eastern Asia was dominant in Asia in terms of patents granted and patents in force in both 2010 and 2020
- China was the dominant country in terms of patents granted in 2020, replacing Japan in 2010
- Asia was the dominant region in terms of patents in force in 2020, replacing Europe in 2010

The trends indicate Eastern Asia, and China in particular, has become dominant in patent activity, suggesting a strong and growing regional investment in research and development (R&D).