Data Wrangling

Data Wrangling

This document showcases how the datasets are wrangled to improve the depth of the variables.

Load tidyverse

The tidyverse package was loaded. This package will make it easy to wrangle and clean data.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Load datasets

The needed datasets are loaded in the column below.

```
econ_free <- read.csv("~/Global-Analysis/Economic_Freedom_data.csv")
df_gap <- read.csv("~/Global-Analysis/new_gapminder.csv")
```

Combining Datasets

Transforming the gapminder dataset

The cell below shows how gapminder was transformed to create a new column.This column will be used to uniquely identify each country during each year.

```
df_gap$char <- as.character(df_gap$year)
df_gap$newcountry <- df_gap$country
df_gap <- df_gap %>%
  unite("country_year", c(newcountry, char), sep = ", ")
```

Checking out gapminder dataset

```
head(df_gap)
```

```
##               country year Life_expectancy Income Population    Continent
## 1         Afghanistan 1800            28.2    603    3280000         Asia
## 2             Albania 1800            35.4    667     400000       Europe
## 3             Algeria 1800            28.8    715    2500000       Africa
## 4             Andorra 1800              NA   1200       2650       Europe
## 5              Angola 1800            27.0    618    1570000       Africa
## 6 Antigua and Barbuda 1800            33.5    757      37000 The Americas
##                country_year
## 1         Afghanistan, 1800
## 2             Albania, 1800
## 3             Algeria, 1800
## 4             Andorra, 1800
## 5              Angola, 1800
## 6 Antigua and Barbuda, 1800
```

Transforming the economic indicators index dataset

The cell below shows how economic indicators dataset was transformed to create a new column.This column will be used to uniquely identify each country during each year.

```
econ_free$char <- as.character(econ_free$Index.Year)
econ_free$newcountry <- econ_free$Name
econ_free <- econ_free %>%
  unite("country_year", c(newcountry, char), sep = ", ")
```

Checking out the dataset

```
head(econ_free)
```

```
##          Name Index.Year Overall.Score Property.Rights Government.Integrity
## 1 Afghanistan       2021            53            30.3                 29.1
## 2     Albania       2021          65.2            46.1                 40.6
## 3     Algeria       2021          49.7              34                 32.7
## 4      Angola       2021          54.2            30.3                 20.4
## 5   Argentina       2021          52.7            46.1                   54
## 6     Armenia       2021          71.9            57.3                   45
##   Judicial.Effectiveness Tax.Burden Government.Spending Fiscal.Health
## 1                   25.7       91.1                76.1          99.9
## 2                   22.8         89                74.6          86.6
## 3                   41.6       67.2                55.4          49.1
## 4                   22.8       87.3                86.9          77.9
## 5                   45.7       70.4                52.8          38.4
## 6                   55.3       87.1                81.3          84.3
##   Business.Freedom Labor.Freedom Monetary.Freedom Trade.Freedom
## 1             53.9          59.9             80.8          68.6
```

```
## 2             66.1          51.6              82          82.8
## 3             63.5          51.3            84.3          57.4
## 4             56.9          59.6            67.5          70.2
## 5             59.5          46.3            41.9          62.6
## 6             81.9          74.5            76.9          73.8
##   Investment.Freedom Financial.Freedom      country_year
## 1                 10                10 Afghanistan, 2021
## 2                 70                70     Albania, 2021
## 3                 30                30     Algeria, 2021
## 4                 30                40      Angola, 2021
## 5                 55                60   Argentina, 2021
## 6                 75                70     Armenia, 2021
```

Joining the two datasets

The cell below was used to join the two columns using the newly formed identifier country_year.

```
?full_join
```

```
## starting httpd help server ... done
```

```
df_world <- inner_join(df_gap, econ_free, by = "country_year")
df_world <- df_world %>%
  select(-c("Name", Index.Year))
```

Saving the datasets

```
write.csv(df_world, "~/Global-Analysis/New_world.csv")
```