

Poster Presentation

Deliverables

The Poster Fair

The Poster

The Oral Presentation

Teamwork

Evaluation

Data Analysis Expectations

Final Project Questions

# STA130H1 Final Project - Poster Presentation Information

*Fall 2018*

## Poster Presentation

The final project for this course will be a statistical analysis of data from The World Factbook (<https://www.cia.gov/library/publications/resources/the-world-factbook/>) on the topic below. You will present your findings in the style of a poster display of a professional scientific conference. You and your team will post your work on a board, give a short oral presentation of your findings when visited by evaluators, and be prepared to answer questions about your work.

## Deliverables

1. A poster to be displayed at the STA130H1 poster fair. See more details about the poster and poster fair below.
2. The R markdown document (.Rmd) that produced the poster and the html (.html) output file, to be submitted before 9:30 on Thursday, December 6 if you are in section LEC0101 and before 13:30 on Thursday, December 6 if you are in section LEC0201. Submission details to follow.
3. A 5 minute oral presentation summarizing your work that you will give at the poster fair.

## The Poster Fair

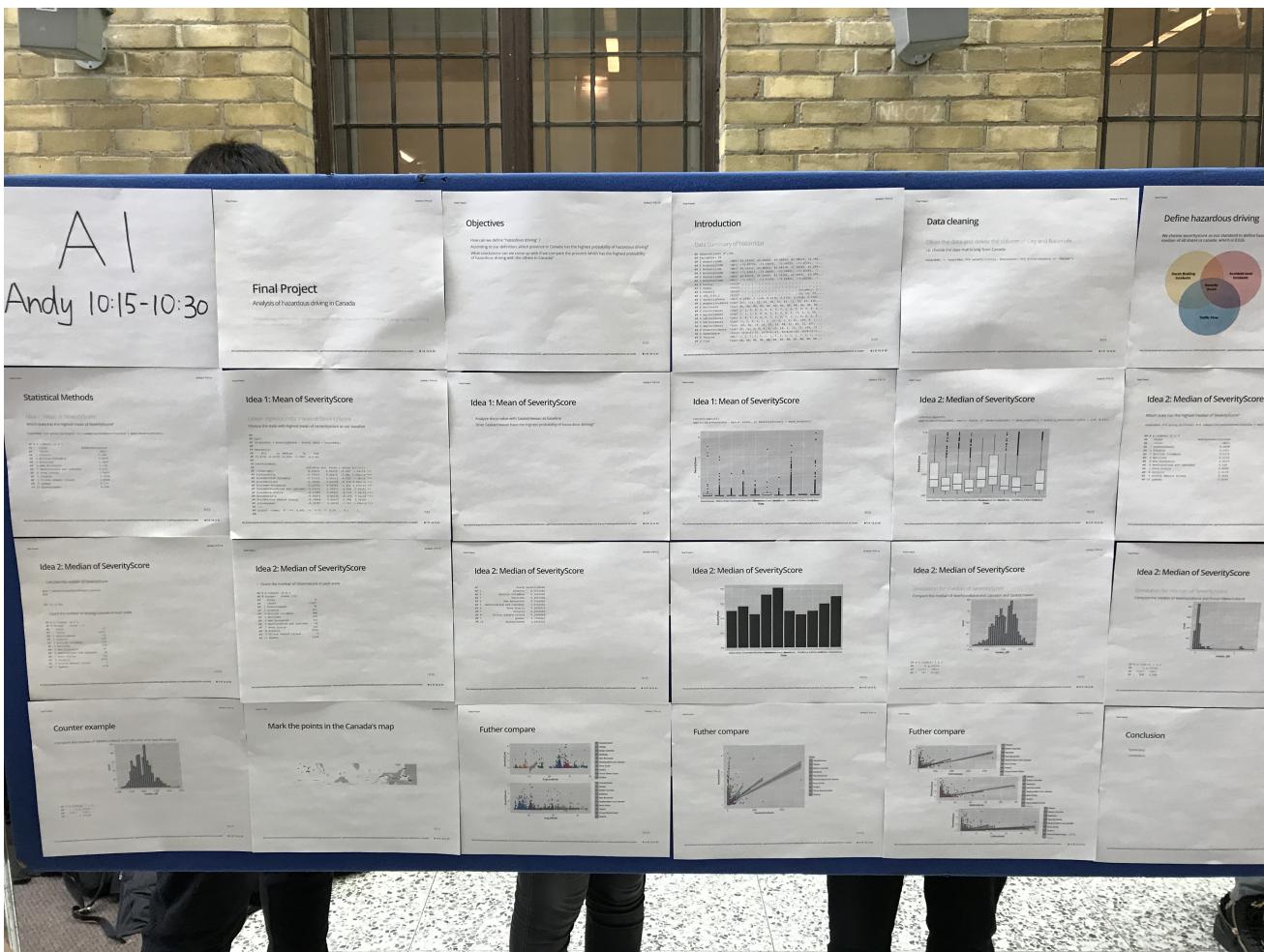
### When

- **Date:** December 6, 2018
- **Time:** During the lecture time that you are registered. If you are registered in LEC0101 then the time is 10:10 - 12:00, or if you are registered in LEC0201 then the time is 14:10 - 16:00. Make sure to come in time to post your work before 10 minutes past the hour. Take down your poster at the end of your class time.

### Location

- Atrium of Bahen Centre, 40 St George Street  
(Note: come to the Bahen atrium and not your lecture room on December 6.)

## The Poster



STA130 Poster Example

- The poster boards on which you will display your work are 6 feet wide by 3 feet high. You must ensure that what your group decides to post will fit on one board.
- Groups can attach individual sheets of paper to the poster board using the velcro that will be provided. No other methods of attachment to the board are permitted.
- You should produce your poster using R markdown to create slides. We recommend ioslides. You can then print the slides on 8.5 x 11 inch sheets of paper, to be posted on the poster board. The skeleton of an R markdown document to produce the slides for your poster is here ([../../project/sta130\\_project\\_template.Rmd](#)). When planning your pages, think of your poster as a grid of these pages, which you will arrange in columns, to effectively show your work.

For more details on using R markdown see:

- The course guide to producing R markdown documents ([../../Rmarkdownforclassreports.html](#))
- R markdown cheat sheet (<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>)
- Information specific to `sta130_project_template.Rmd` ioslides ([https://rmarkdown.rstudio.com/ioslides\\_presentation\\_format.html](#))
- Your poster should include R output, but not R code. See the template ([../../project/sta130\\_project\\_template.Rmd](#)) for an example of how to write code chunks to do this.
- Your poster must include the names of your team members, your tutorial section (e.g., TUT0101A), and your group number as assigned by your TA.
- Your poster should include the following sections: Introduction, Statistical Methods, Results, Conclusion. See the template ([../../project/sta130\\_project\\_template.Rmd](#)) for optional sections.
- See the Self-Evaluation Checklist for STA130 Project ([../../project/STA130\\_project\\_checklist.pdf](#)) and the template R markdown document ([../../project/sta130\\_project\\_template.Rmd](#)) for details on what to include.

A few guidelines for an effective poster display:

- Your poster should be clear, concise, and easy to read quickly.
- Do not use small fonts as it will need to be read at a distance of about 6 feet.
- Figures display information more efficiently than text.
- Numbered or bulleted lists convey points in a poster setting more effectively than blocks of text.

## The Oral Presentation

At the poster fair you will be asked to give a 5 minute presentation summarizing your work. This time limit is firm and you will be asked to stop when time is up. Each team member must speak during this presentation.

## Teamwork

- Preparatory work will be carried out in tutorial and will form part of your tutorial grade.
- You will work in a group comprised of either 3 or 4 students.
- You must work with students in your tutorial only. Some tutorial time will be used for project work.

- Your TA can help you form a group with other students in your tutorial.
- All team members are expected to contribute equally to the completion of the project. All team members must present part of the oral presentation at the poster fair.
- Your group may not work with members of another group. You may not discuss the project with anyone except for your team, professors, and the course TAs.

## Evaluation

Grade component*	Value
Poster Content	50%
Reproducibility of Poster Content	10%
Oral Presentation	40%

\* Students will be evaluated as a team.

## Poster Content and Reproducibility

- The rubric for the poster evaluation is here ([..../project/STA130\\_project\\_poster\\_rubric.pdf](#)).
- Before the poster fair, you must send your TA the html and R Markdown files of your poster. These are due at 9:30 if you are in section LEC0101 and at 13:30 if you are in section LEC0201.
- Your TA will attempt to reproduce your poster using the R Markdown (.rmd) files you submit.
- If your TA cannot run the .rmd files you submit to reproduce your poster content then your group will receive 0; if the TA has to make minor changes to get it to run then your group will receive 1; and if it runs with no changes then your group will receive 2.
- If the R Markdown and HTML files are submitted after the deadline every member of your group will lose 10% of your overall final project mark as long as they are submitted at most 24 hours after the deadline. The R Markdown and HTML files will not be accepted more than 24 hours after the deadline.

## Oral Presentation

- During the poster fair, you will be visited by members of the STA130H1 teaching team. You will give them a 5 minute presentation about your work. The rubric for the oral presentation is here ([..../project/STA130\\_project\\_oral\\_presentation\\_rubric.pdf](#)).
- Every member of the team is expected to speak as part of the oral presentation.
- If a student in a group isn't present at their group's presentation then they will need a valid excuse (e.g., UoT illness form), otherwise they will receive 50% of the group mark.
- If a student doesn't speak at all during the presentation and is unable to answer a direct question then they will receive 50% of the group mark. If a student neither speaks nor responds to any questions they will receive 0.

## Data Analysis Expectations

You will carry out a data analysis on data from the country comparisons pages of The World Factbook (<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/rankorderguide.html>) using R to address the topic below.

We expect that your analysis will require data wrangling, exploratory data analysis (plots and summary statistics), tests, confidence intervals, classification trees, and regression models. Your project does not need to include all of these statistical methods nor does it need to include all of the variables in the data sets. You might also choose not to include all observations, or to make new variables from the data that may be more suitable for answering your questions of interest.

The goal is not to carry out an exhaustive analysis, nor to apply everything you have learned in the course. The goal is to demonstrate that you have learned how to use R, that you can appropriately apply the methods we have covered in class to address a question, and that you can effectively interpret and present the results.

## The Data

The map below shows the number of global internet users according to the CIA WorldFactbook.

```

library(tidyverse)
library(maps)

world <- map_data("world")

internetusers_cia2017 <- read_csv("internetusers_cia2017.csv")

iu <- internetusers_cia2017 %>% rename(region = Country)

iu$region[4] <- "USA" # to match world map data

iu <- semi_join(iu, world, by = "region") #only keep countries according to world map data

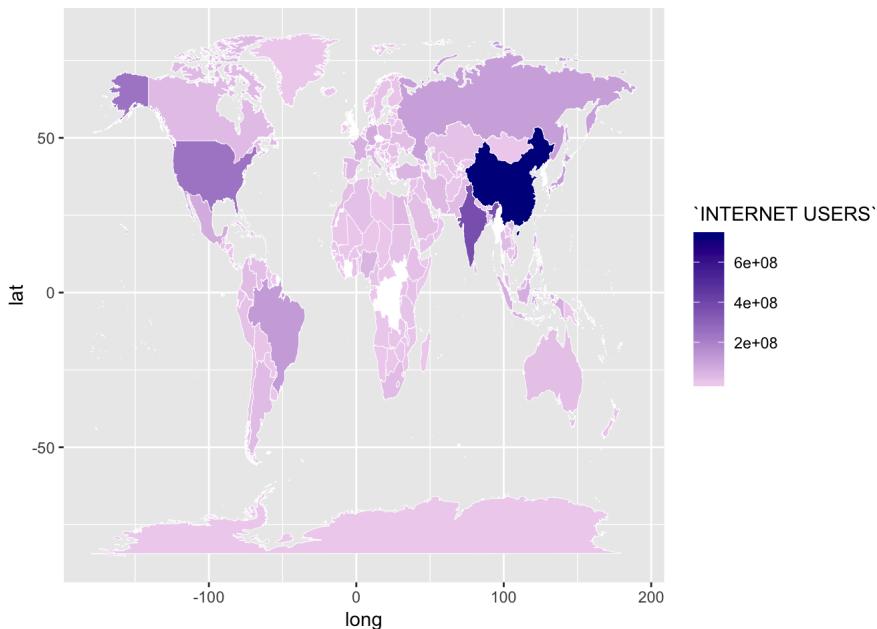
# code below is modified from
# https://stackoverflow.com/questions/29614972/ggplot-us-state-map-colors-are-fine-polygons-jagged-r
gg <- ggplot()

gg <- gg + geom_map(
  data = world,
  map = world,
  aes(x = long, y = lat, map_id = region),
  fill = "#ffffff",
  color = "#ffffff",
  size = 0.20
)

gg <- gg + geom_map(
  data = iu,
  map = world,
  aes(fill = `INTERNET USERS`, map_id = region),
  color = "#ffffff",
  size = 0.15
)

gg <- gg + scale_fill_continuous(low = 'thistle2', high = 'darkblue',
guide = 'colorbar')
gg

```



Your group will explore several data sets from the CIA World Factbook. We have proposed questions for your group to address in your project. There are many ways you can address these questions in the data. Your group will need to focus your project, choosing what you will consider. You do not need to consider every variable in the data set.

The country comparison data sets are described here (<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/rankorderguide.html>). The World Factbook website contains information on all the data. For example, the References and Notes section (<https://www.cia.gov/library/publications/resources/the-world-factbook/docs/notesanddefs.html>) contains data definitions.

## Frequently Asked Questions about the Data

Coming soon ...

## Final Project Questions

"The Internet is the decisive technology of the Information Age, and with the explosion of wireless communication in the early twenty-first century, we can say that humankind is now almost entirely connected, albeit with great levels of inequality in bandwidth, efficiency, and price." according to an article in the MIT Technology Review (<https://www.technologyreview.com/s/530566/the-impact-of-the-internet-on-society-a-global-perspective/>) by Manuel Castells. Most people think that internet use is a good thing for individuals and society (for example see here (<http://www.pewinternet.org/2018/04/30/declining-majority-of-online-adults-say-the-internet-has-been-good-for-society/>) and here (<http://www.pewglobal.org/2018/10/09/internet-connectivity-seen-as-having-positive-impact-on-life-in-sub-saharan-africa/>)). Yet there is a global digital divide ([https://en.wikipedia.org/wiki/Digital\\_divide](https://en.wikipedia.org/wiki/Digital_divide)) with regard to internet use. Yet, another article (<https://www.technologyreview.com/the-download/608887/the-un-says-the-global-digital-divide-could-become-a-yawning-chasm/>) states that "more than 52 percent of people on the planet still don't have Internet access. Men outnumber women as Web users in every region of the world. And there remain massive disparities in connection speeds in different countries. These are just some of the major findings outlined in a new United Nations ([https://www.itu.int/dms\\_pub/itu-s/obp/pol/S-POL-BROADBAND.18-2017-PDF-E.pdf](https://www.itu.int/dms_pub/itu-s/obp/pol/S-POL-BROADBAND.18-2017-PDF-E.pdf)) report about the state of the world's Internet connections."

## Data

Use the following data sets from the CIA World Factbook and the democracy index ([http://pages.eiu.com/rs/753-RIQ-438/images/Democracy\\_Index\\_2017.pdf?](http://pages.eiu.com/rs/753-RIQ-438/images/Democracy_Index_2017.pdf?mkt_tok=eyJpIjoiWkRKbU1HWmxNVEUwTW1FdylsInQiOjPdltVFV0bIFRQzZNVERCZHhVeitZRElmUGpIOHh3NWs1d2wzVzdRS1JvNU1kVmUxQVRESL)

(<http://www.eiu.com/home.aspx>) developed by the Economist Intelligence Unit (<http://www.eiu.com/home.aspx>)

### Using rstudio.cloud

[https://en.wikipedia.org/wiki/Democracy\\_Index#Democracy\\_Index\\_by\\_country\\_\(2017\)](https://en.wikipedia.org/wiki/Democracy_Index#Democracy_Index_by_country_(2017))  
([https://en.wikipedia.org/wiki/Democracy\\_Index#Democracy\\_Index\\_by\\_country\\_\(2017\)](https://en.wikipedia.org/wiki/Democracy_Index#Democracy_Index_by_country_(2017)))

```
democracyindex2017 <- read_csv("democracyindex2017.csv")
democracyindex2017 %>% head()
```

Rank	Country	Score	Electoral process and functioning of government	Political participation	
1	Norway	9.87	10.00	9.64	10.00
2	Iceland	9.58	10.00	9.29	8.89
3	Sweden	9.39	9.58	9.64	8.33
4	New Zealand	9.26	10.00	9.29	8.89
5	Denmark	9.22	10.00	9.29	8.33
6	Ireland	9.15	9.58	7.86	8.33

<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2206rank.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2206rank.html>)

```
education_cia2017 <- read_csv("education_cia2017.csv")
education_cia2017 %>% head()
```

Rank	Country	Expenditure on education (% of GDP)	Date of information
1	Lesotho	13	2008
2	Cuba	12.8	2010
3	Marshall Islands	12.2	2003
4	Kiribati	12	2001
5	Botswana	9.5	2009
6	Sao Tome and Principe	9.5	2010

<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2004rank.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2004rank.html>)

```
gdppp_cia2017 <- read_csv("gdppp_cia2017.csv")
gdppp_cia2017 %>% head()
```

Rank	Country	GDP - PER CAPITA (PPP)	Date of Information
1	Liechtenstein	\$139,100	2009 est.
2	Qatar	\$124,500	2017 est.
3	Monaco	\$115,700	2015 est.
4	Macau	\$111,600	2017 est.
5	Luxembourg	\$106,300	2017 est.
6	Bermuda	\$99,400	2016 est.

<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2102rank.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2102rank.html>)

```
lifeexpect_cia2017 <- read_csv("lifeexpect_cia2017.csv")
lifeexpect_cia2017 %>% head()
```

```
## # A tibble: 6 x 4
##   Rank Country     `(YEARS)` `Date of Information`
##   <int> <chr>      <dbl> <chr>
## 1     1 Monaco      89.4 2017 est.
## 2     2 Japan        85.3 2017 est.
## 3     3 Singapore    85.2 2017 est.
## 4     4 Macau        84.6 2017 est.
## 5     5 San Marino   83.3 2017 est.
## 6     6 Iceland      83.1 2017 est.
```

<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2225rank.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2225rank.html>)

```
healthxpend_cia2017 <- read_csv("healthxpend_cia2017.csv")
healthxpend_cia2017 %>% head()
```

```
## # A tibble: 6 x 4
##   Rank Country     `(% OF GDP)` `Date of Information`
##   <int> <chr>      <dbl> <int>
## 1     1 United States 17.1  2014
## 2     2 Marshall Islands 17.1  2014
## 3     3 Tuvalu        16.5  2014
## 4     4 Maldives       13.7  2014
## 5     5 Micronesia, Federated States of 13.7  2014
## 6     6 Sweden         11.9  2014
```

<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2153rank.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2153rank.html>)

```
internetusers_cia2017 <- read_csv("internetusers_cia2017.csv")
internetusers_cia2017 %>% head()
```

```
## # A tibble: 6 x 4
##   Rank Country     `INTERNET USERS` `Date of Information`
##   <int> <chr>      <dbl> <chr>
## 1     1 China        730723960 July 2016 est.
## 2     2 European Union 398100000 July 2016 est.
## 3     3 India         374328160 July 2016 est.
## 4     4 United States 246809221 July 2016 est.
## 5     5 Brazil         122841218 July 2016 est.
## 6     6 Japan          116565962 July 2016 est.
```

<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2150rank.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2150rank.html>)

```
telephonelines_cia2017 <- read_csv("telephonelines_cia2017.csv")
telephonelines_cia2017 %>% head()
```

```
## # A tibble: 6 x 4
##   Rank Country     `TELEPHONES - MAIN LINES IN USE` `Date of Informati...
##   <int> <chr>      <dbl> <chr>
## 1     1 China        193762000 2017 est.
## 2     2 United States 119902000 2017 est.
## 3     3 Japan         63941094 2017 est.
## 4     4 Germany       44400000 2017 est.
## 5     5 Brazil         40878018 2017 est.
## 6     6 France        38687000 2017 est.
```

<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2119rank.html>  
(<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2119rank.html>)

```
population_cia2017 <- read_csv("population_cia2017.csv")
population_cia2017 %>% head()
```

```

## # A tibble: 6 x 4
##   Rank Country      POPULATION `Date of Information`
##   <int> <chr>          <dbl> <chr>
## 1     1 China           1379302771 July 2017 est.
## 2     2 India            1281935911 July 2017 est.
## 3     3 United States    326625791 July 2017 est.
## 4     4 Indonesia        260580739 July 2017 est.
## 5     5 Brazil            207353391 July 2017 est.
## 6     6 Pakistan          204924861 <NA>

```

[https://meta.wikimedia.org/wiki/List\\_of\\_countries\\_byRegional\\_classification](https://meta.wikimedia.org/wiki/List_of_countries_byRegional_classification)  
[\(https://meta.wikimedia.org/wiki/List\\_of\\_countries\\_byRegional\\_classification\)](https://meta.wikimedia.org/wiki/List_of_countries_byRegional_classification)

```

world_regions <- read_csv("world_regions.csv")
world_regions %>% head()

```

```

## # A tibble: 6 x 3
##   Country       Region      `Global South`
##   <chr>         <chr>       <chr>
## 1 Andorra       Europe      Global North
## 2 United Arab Emirates Arab States Global South
## 3 Afghanistan   Asia & Pacific Global South
## 4 Antigua and Barbuda South/Latin America Global South
## 5 Anguilla      South/Latin America Global South
## 6 Albania       Europe      Global North

```

## Using RStudio on your own computer

Copy, paste, and run the code chunk below to download all of the data into RStudio on your own computer.

```

library(tidyverse)
path <- "https://raw.githubusercontent.com/ntaback/UofT_STA130/master/Fall2018/Finalproject/"

democracyindex2017 <- read_csv(paste0(path, "democracyindex2017.csv"))
education_cia2017 <- read_csv(paste0(path, "education_cia2017.csv"))
gdppp_cia2017 <- read_csv(paste0(path, "gdppp_cia2017.csv"))
lifeexpect_cia2017 <- read_csv(paste0(path, "lifeexpect_cia2017.csv"))
healthxpend_cia2017 <- read_csv(paste0(path, "healthxpend_cia2017.csv"))
internetusers_cia2017 <- read_csv(paste0(path, "internetusers_cia2017.csv"))
telephonelines_cia2017 <- read_csv(paste0(path, "telephonelines_cia2017.csv"))
population_cia2017 <- read_csv(paste0(path, "population_cia2017.csv"))
world_regions <- read_csv(paste0(path, "world_regions.csv"))

```

## Accessing Other Data from the World Factbook

There are many other data sets (<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/rankorderguide.html>) available from The World Factbook. An R function `get_CIAWFB_data` has been written to allow you to access these data sets easily.

```

get_CIAWFB_data <- function(table_url) {
  library(rvest)
  dat <- xml2::read_html(table_url) %>% rvest::html_table()
  dat[[1]]
}

```

Suppose your group is considering using the median age country comparison data as supplemental data. The url for this table is <https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2177rank.html> (<https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2177rank.html>). One way to easily import this data into R is by using `get_CIAWFB_data`.

```

medianage_cia2017 <- get_CIAWFB_data("https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2177rank.html")
medianage_cia2017 %>% head()

```

```

##   Rank          Country MEDIAN AGE Date of Information
## 1     1           Monaco      53.1    2017 est.
## 2     2           Japan       47.3    2017 est.
## 3     3          Germany      47.1    2017 est.
## 4     4 Saint Pierre and Miquelon    46.5    2017 est.
## 5     5           Italy       45.5    2017 est.
## 6     6          Slovenia      44.5    2017 est.

```

## Other Sources of Supplemental Data

Another source of Global data that can be used to supplement the data you are given is from the United Nations (<http://data.un.org/Explorer.aspx?d=UNODC>).

## Questions

1. Define how you will measure internet use in a country. Note that we expect that each group may have a different definition. The analysis your group does will follow from how you choose to define internet use in a country.
2. Do different regions of the world have different internet usage?
3. According to your definition of internet use what is the impact of democracy, education, economy, and health on internet use? Your group will need to define democracy, education, economy, and health.

To answer questions 1 - 3, you must use the data sets in the section above, although you don't have to use all the data. It is not necessary to use data from any other source. However you are allowed to supplement the data provided with other data if you'd like.