Population Project

Boon Yeow Koh and Anuj Patel

Problem and Background:

Population growth is a large problem that is facing the world today, with global populations sky rocketing since 1960. This is a really interesting situation because there are so many different possible reasons for this happening, from social changes to environmental changes. We decided to focus on food and water availability because historically those are the two largest building blocks of civilizations. Food growth is slated to be a large problem for the future, with the United Nation's Food and Agriculture Organization estimating that a 70% increase in food production is needed by 2050. It is critical to understand what is causing population growth so that it is possible to keep up with the growing demand for resources around the world. Countries that suffer from water poverty also have higher birth rates, and quickly growing populations, with almost double the normal birth rate. This is problematic for the future as a greater population with less water availability is a cause for concern.

Methods and Results:

The first thing that we did was find population figures over time for the past 50 years, from 1961 to 2013. This is a very representative of the time when population growth really started to increase. We then looked at calorie and water availability per capita for each country and graphed them against population growth. This gave us an initial idea of how the relationship changed over time for many different countries. The next thing we did was look at the relationship for developing nations, because this is where most of the population growth has happened over our chosen time period. This was really interesting as even there only China and India had very large population growths, while most other nations stayed relatively stable. Finally we looked at it from the continents point of view, as certain continents such as Asia and Africa have seen much larger growth than North America or Europe.

Results and Implications:

The most important thing that we found is that there seems to be diminishing returns to the amount of calories available per capita as it relates to population growth. After about 2600-2800 calories, population growth seems to slow down across the board. However, this trend does not work for China which seems to be an outlier. It started with very high population and low calories, around 1500 per day per capita, but is still only growing from there. For the future, it would be very important to look at social factors that could lead to population growth in places like this, because it does not seem like calories are the major contributing factor here. Food distribution could also be another large factor, as the calories may be available but possible not available to the people that need them. Income did not have any effect on potable drinking water, however drinking water was a necessity. However, after 65% of the population had access to potable water, the relationship with population was not as strong. Again, it is important to look for other outside factors that could be in effect here.

Data wrangling, munging and cleaning

We Downloaded the a CSV file for Food Supply from Our World Data. Our World Data gets their data from United Nations Food and Agriculture Organization (UNFAO). For the water CSV file, we only need the data for the safetly managed drinking water, so we deleted everything else that we did not need. The water CSV file did not include total population data for each country, so we added the population data into the file. In order to create the food graphs, we had to filter the food data to 1961-2013 as there is missing data for food supply before 1961. For both the food and water graphs, we had to filter out data to find data for continents, income, individual countries and cummulative world data. The code to filter and clean data can be found in the appendix.

References

Food Data: Roser, Max, and Hannah Ritchie. "Food Supply." Our World in Data, 5 Mar. 2013, https://ourworldindata.org/food-supply.

Water Data: "Water, Sanitation and Hygiene (WASH) Data Explorer." Our World in Data, https://ourworldindata.org/explorers/water-and-sanitation.

Appendix

Data Cleaning and Filtering for Food

Filtering out all of the data for continent just to get data for individual countries

```
food = read.csv("global-food-vs-life.csv")

foodCountries = food %>% filter(Year > 1960 & Year < 2014 & Country != "Africa" & Country
  != "Eastern Asia" & Country != "Eastern Europe" & Country != "Asia" & Country !=
  "Asia, Central" & Country != "Oceania" & Country != "No income group available" &
  Country != "Europe and Northern America" & Country != "Europe and Northern America" &
  Country != "Europe" & Country != "Europe, Western" & Country !=
  "Least Developed Countries" & Country != "Latin America and the Caribbean" & Country !=
  "North America" & Country != "South America" & Country != "World")
```

Filtering to get data for all continents

```
foodContinent = food %>% filter(Year > 1960 & Year < 2014 & (Country == "Africa" | Country
  == "Asia" | Country == "Oceania" | Country == "Europe" | Country == "South America"))
```

Global Food Supply Animation from 1961 to 2013

This graph shows a time animation of the relationship between daily calories per capita and population for all countries from 1961 to 2013. The Animation code is commented becasuse it will produce too many imaged inside the PDF.

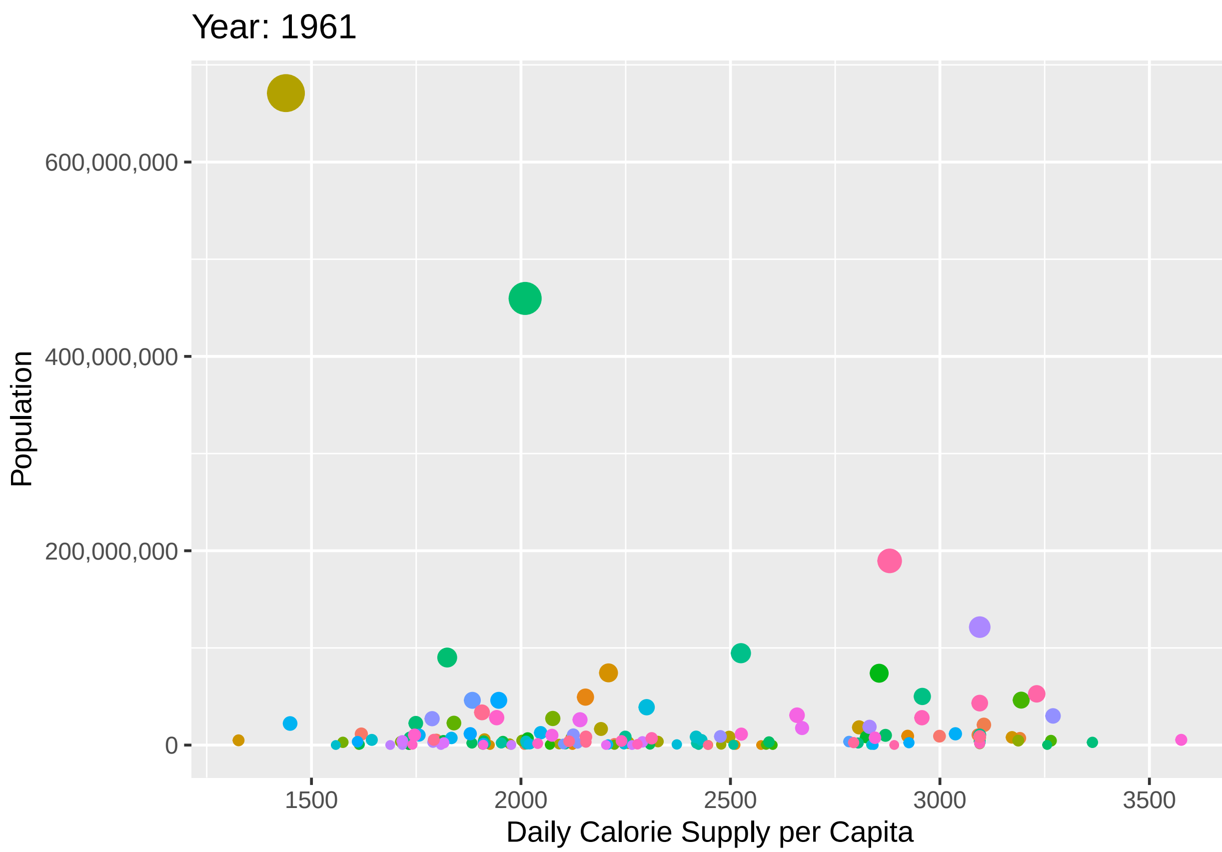
```
g = ggplot(foodCountries, aes(x = Calories, y=Population, size = Population, color = Country)) +
  geom_point(show.legend = FALSE, alpha = 0.7) +
  scale_size(range = c(2, 12)) +
  scale_x_log10() +
  labs(title = 'Year: {frame_time}', x = "Daily Calorie Supply per Capita", y = "Population") +
```

```
  scale_y_continuous(labels = comma) # +
  # transition_time(Year) +
  # ease_aes('linear')

# animate(g, duration = 10, fps = 20, width = 500, height = 500, renderer = gifski_renderer())
```

Global Food Supply in 1961

This graph shows the relationship between daily calories per capita and population in 1961 for all countries.

```
food1961 = foodCountries %>% filter(Year == 1961) #get data for 1961

ggplot(food1961, aes(x=Calories, y=Population, color = Country)) + geom_point(show.legend = FALSE,
  aes(size = Population)) + labs(title = 'Year: 1961', x = "Daily Calorie Supply per Capita", y =
  "Population") + scale_y_continuous(labels = comma) #display scatter plot
```

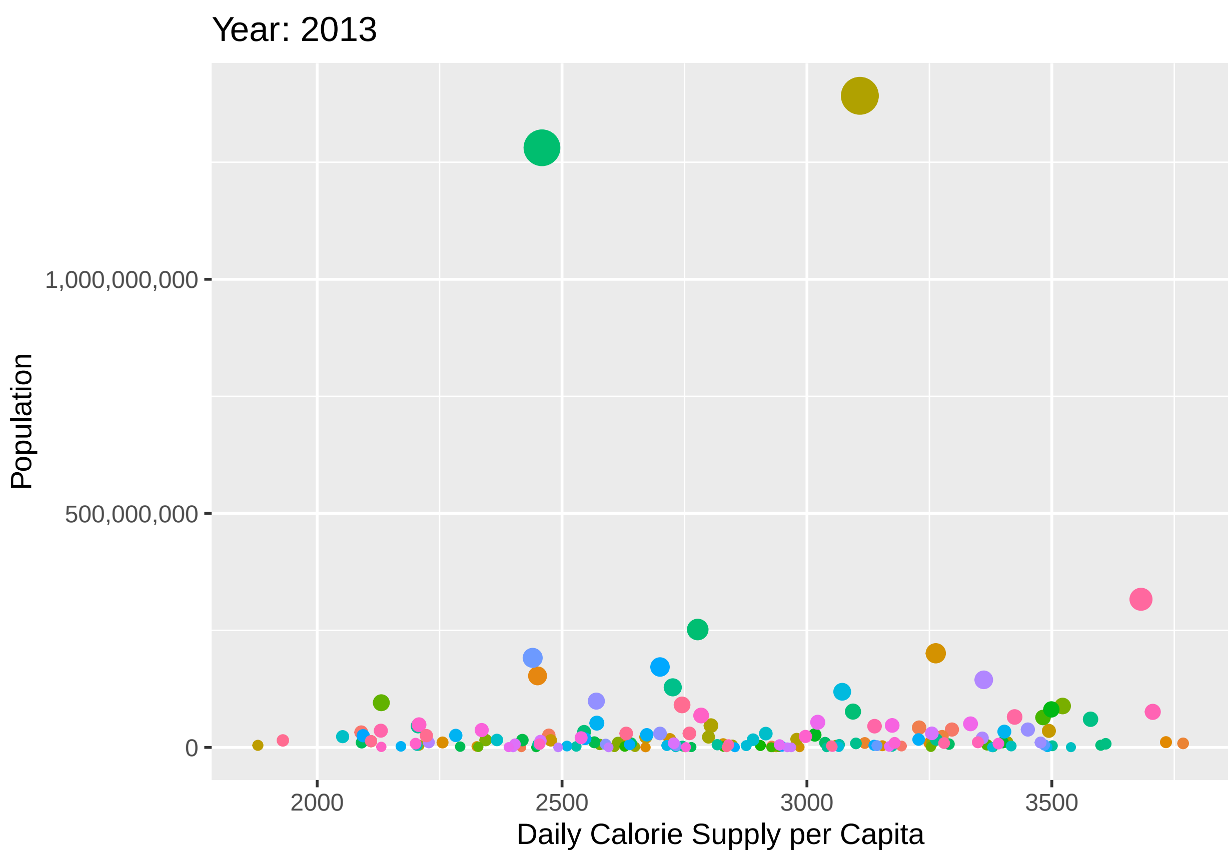


Global Food Supply in 2013

This graph shows the relationship between daily calories per capita and population in 2013 for all countries.

```
food2013 = foodCountries %>% filter(Year == 2013) #get data for 2013
```

```r
ggplot(food2013, aes(x=Calories, y=Population, color = Country)) + geom_point(show.legend =
  FALSE, aes(size = Population)) + labs(title = 'Year: 2013', x =
  "Daily Calorie Supply per Capita", y = "Population") + scale_y_continuous(labels = comma)
```

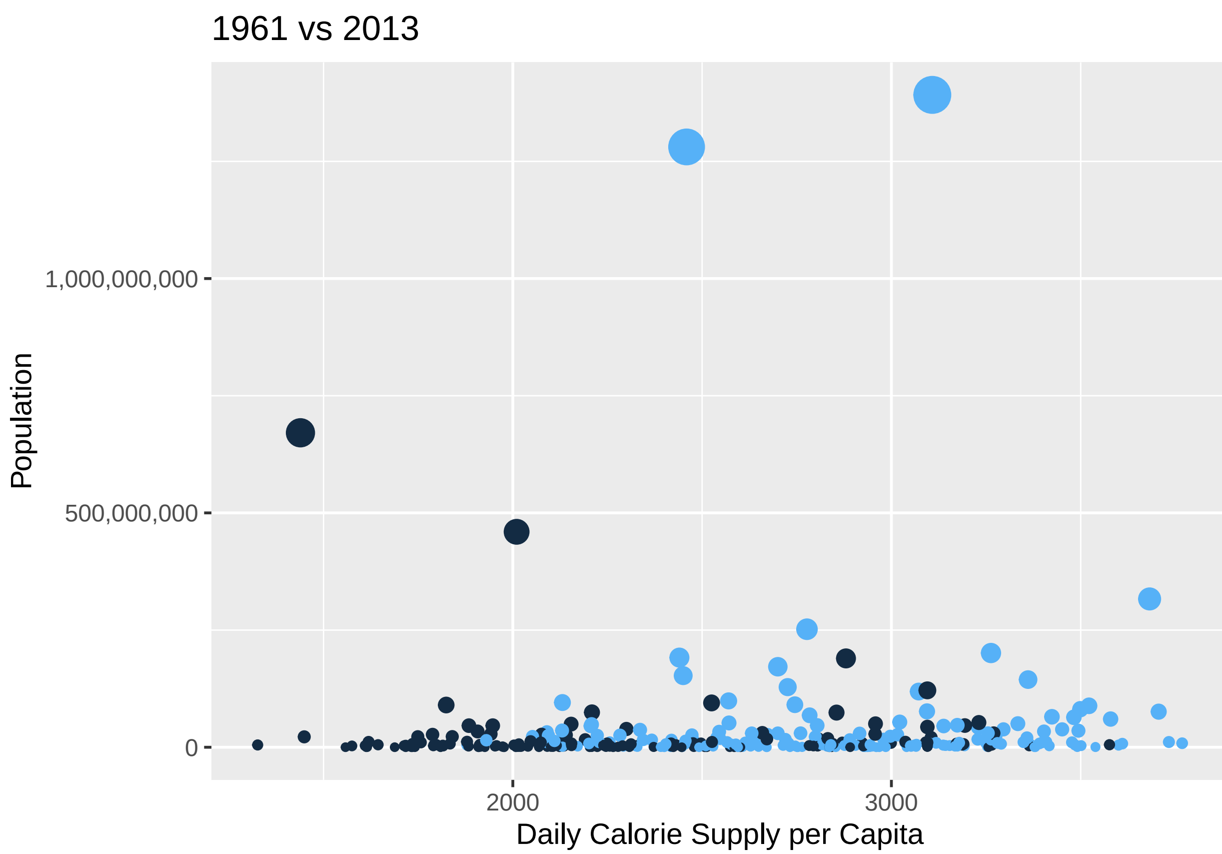



```r
  #display scatter plot
```

Global Food Supply Comparison between 1961 and 2013

This graph shows the comparison of 1961 and 2013 for all countries for the relationship between daily calories per capita and population.

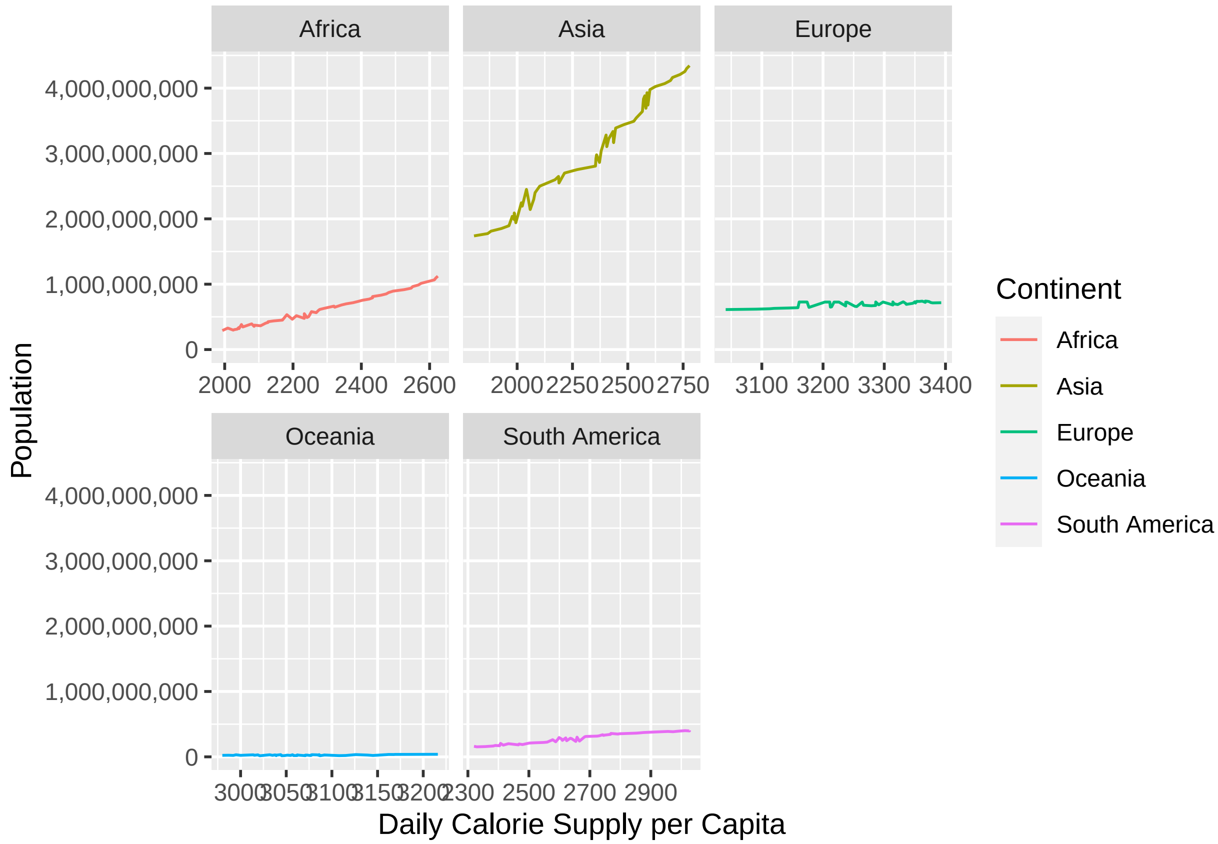
```r
#get data for 1961 and 2013
food1961_2013 = foodCountries %>% filter(Year == 2013 | Year == 1961)

#display scatter plot
ggplot(food1961_2013, aes(x=Calories, y=Population)) + geom_point(show.legend = FALSE,
  aes(size = Population, color = Year)) + labs(title = '1961 vs 2013', x =
  "Daily Calorie Supply per Capita", y = "Population") + scale_y_continuous(labels = comma)
```



Food Supply of Various Continents from 1961-2013

This graph shows the relationship between Daily Calories per capita and population of different continents from 1961-2013

```
ggplot(data =foodContinent, mapping = aes(x = Calories, y = Population, color=Country)) +
  geom_line()+facet_wrap(~Country, scales="free_x") + labs(fill = "Continent", x =
  "Daily Calorie Supply per Capita", y = "Population") + scale_y_continuous(labels = comma) +
  guides(color=guide_legend("Continent")) #display scatter plot
```



Data Cleaning and Filtering for Water

Get data for individual countries and filtering out income and region data

```
waterData = read.csv("water.csv")

waterCountries = waterData %>% filter(Country != "Low income" & Country !=
  "Lower-middle income" & Country != "Landlocked Developing Countries" & Country !=
  "Fragile or Extremely Fragile" & Country != "High income" & Country !=
  "Least Developed Countries" & Country != "Upper-middle income" & Country != "World"
  & Country != "Latin America and the Caribbean" & Country !=
  "Western Asia and Northern Africa" & Country != "Eastern and South-Eastern Asia" &
  Country != "North America and Europe" & Country == "Sub-Saharan Africa")

head(waterData)
```

```
##   Normal.Water Clean.Water     Country Year Water.Access  Population
## 1     31.83205    10.92753 Afghanistan 2000      6614687 20779957.03
## 2     31.86091    10.93879 Afghanistan 2001      6884184 21606992.19
## 3     34.18586    11.72537 Afghanistan 2002      7726268 22600773.44
## 4     36.51088    12.51212 Afghanistan 2003      8646094 23680871.09
## 5     38.85347    13.30600 Afghanistan 2004      9607178 24726689.45
## 6     41.21403    14.10727 Afghanistan 2005     10573161 25654273.44
```

Get data for region The data for the safetly managed drinking water for East Asia and South East Asia is not available in the dataset

```
waterRegion = waterData %>% filter(Country == "Latin America and the Caribbean" | Country
  == "Western Asia and Northern Africa" | Country == "North America and Europe" | Country
  == "Sub-Saharan Africa")

head(waterRegion)
```

```
##   Normal.Water Clean.Water                         Country Year Water.Access
## 1     91.42698    71.73844 Latin America and the Caribbean 2000    477099177
## 2     91.81345    72.06466 Latin America and the Caribbean 2001    485888868
## 3     92.19712    72.36624 Latin America and the Caribbean 2002    494583689
## 4     92.57309    72.66216 Latin America and the Caribbean 2003    503175619
## 5     92.93902    72.95019 Latin America and the Caribbean 2004    511680667
## 6     93.29961    73.23480 Latin America and the Caribbean 2005    520146232
##    Population
## 1 521836325.2
## 2 529213152.9
## 3   536441569
## 4 543544139.4
## 5 550555249.6
## 6 557500980.4
```

Get data for income levels

```
waterIncome = waterData %>%  filter(Country == "Low income" | Country == "Lower-middle income"
  | Country == "High income" |Country == "Upper-middle income")

head(waterIncome)
```

```
##   Normal.Water Clean.Water     Country Year Water.Access Population
## 1     98.91557    95.15429 High income 2000   1065136614 1076813916
## 2     98.93314    95.19629 High income 2001   1072024597 1083584943
## 3     99.17216    95.78709 High income 2002   1081418244 1090445370
## 4     99.19069    95.84531 High income 2003   1088615798 1097497967
## 5     99.20956    95.90658 High income 2004   1096140297 1104873658
## 6     99.13565    95.81216 High income 2005   1103028244 1112645379
```

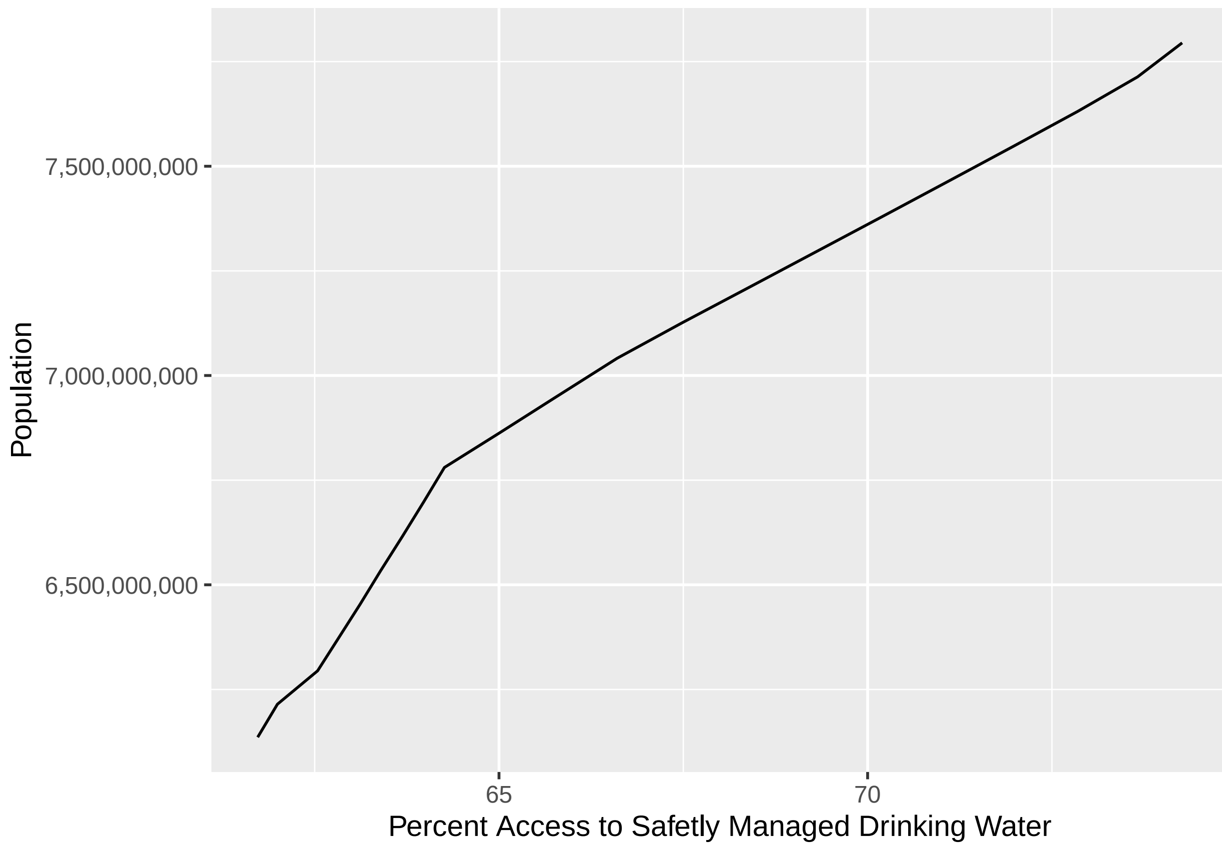
Get Cummulative World Data

```
waterWorld = waterData %>% filter(Country == "World")

head(waterWorld)
```

```
##   Normal.Water Clean.Water Country Year Water.Access Population
## 1     84.19318    61.72681   World 2000   5165922182 6135796410
## 2     84.56640    61.99468   World 2001   5255518179 6214664420
## 3     85.13025    62.53974   World 2002   5358499057 6294471460
## 4     85.65798    62.82667   World 2003   5459527380 6373635310
## 5     86.18056    63.11558   World 2004   5561516825 6453330955
## 6     86.69247    63.39525   World 2005   5664273035 6533754609
```

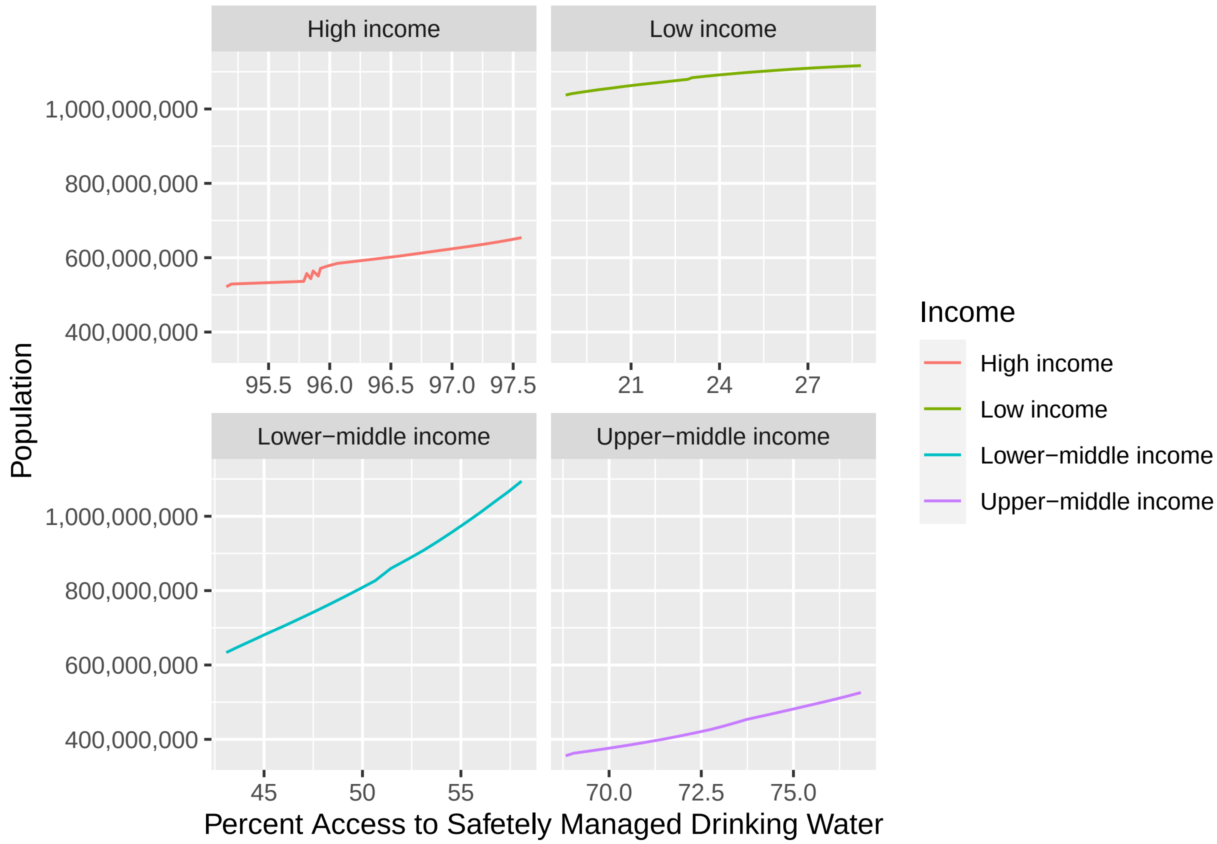
Clean World Water Supply from 1961-2013

```
waterWorld$Population <- as.numeric(waterWorld$Population)

ggplot(data = waterWorld, mapping = aes(x = Clean.Water, y=Population)) + geom_line(aes(group=1)) +
  labs( x = "Percent Access to Safetly Managed Drinking Water", y = "Population") +
  scale_y_continuous(labels = comma)
```



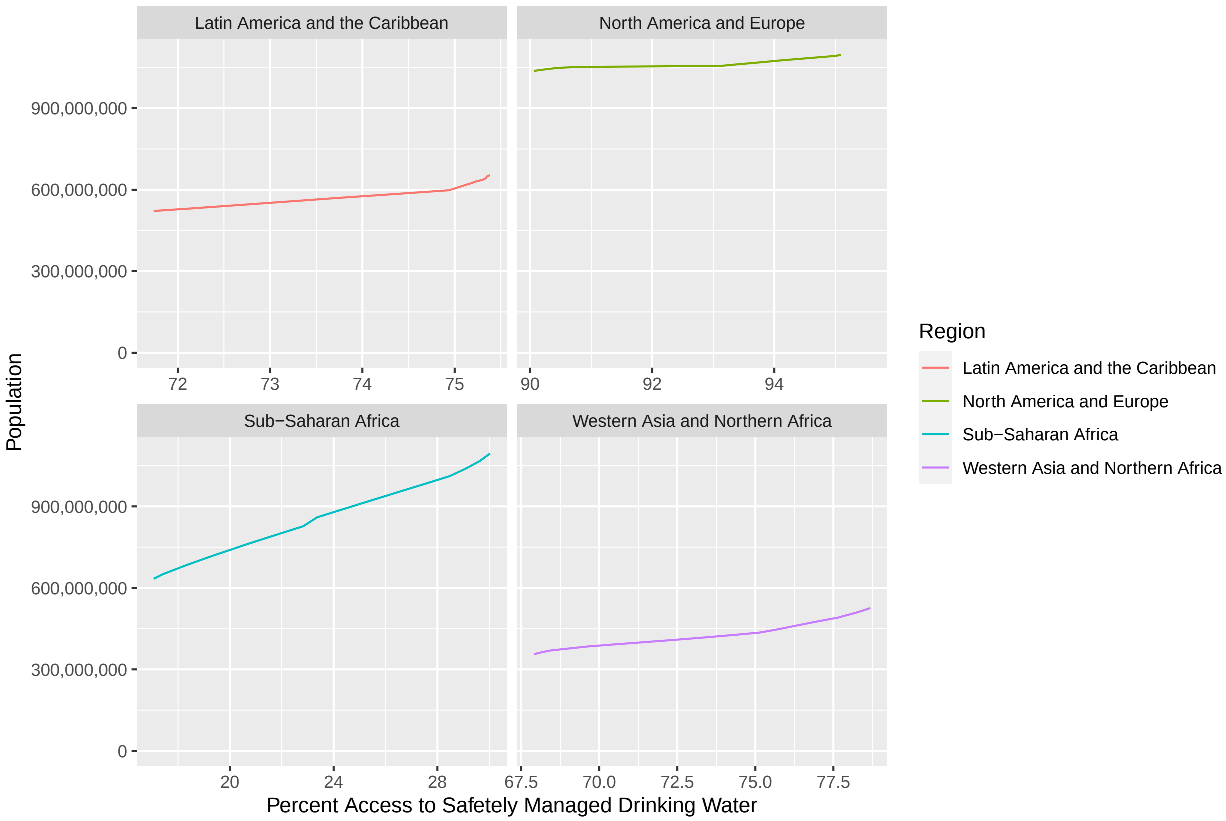
Clean Water Supply of countries grouped by income levels from 1961-2013

```
waterIncome$Population <- as.numeric(waterRegion$Population)

ggplot(data = waterIncome, mapping = aes(x = Clean.Water, y=Population, color = Country)) +
  geom_line(aes(group=Country)) + labs(x = "Percent Access to Safetely Managed Drinking Water",
  y = "Population") + guides(color=guide_legend("Income")) + scale_y_continuous(labels = comma) +
  facet_wrap(~Country, scales="free_x")
```



Clean Water Supply of countries grouped by income levels from 1961-2013

```
waterRegion$Population <- as.numeric(waterRegion$Population)

ggplot(data = waterRegion, mapping = aes(x = Clean.Water, y=Population, color = Country)) +
  geom_line(aes(group=Country)) + labs(x = "Percent Access to Safetely Managed Drinking Water",
  y = "Population") + guides(color=guide_legend("Region")) +
  scale_y_continuous(limits=c(0, 1100000000), labels = comma) +
  facet_wrap(~Country, scales="free_x")
```


Latin America and the Caribbean
North America and Europe
Sub−Saharan Africa
Western Asia and Northern Africa
Population
900,000,000
600,000,000
300,000,000
0
72
73
74
75
90
92
94
20
24
28
67.5
70.0
72.5
75.0
77.5
Percent Access to Safetely Managed Drinking Water
Region
Latin America and the Caribbean
North America and Europe
Sub−Saharan Africa
Western Asia and Northern Africa