

Global Coal Consumption

Jackie Kim

12/14/2018

Contents

Introduction	1
Tidying Data	1
Data Visualization	3
Finding a Reason	4
Visualizing Countries in Asia and Oceania	5
Ratio between China and Asia & Oceania Region	6
China vs the World	7
Conclusion	8

Introduction

This R Markdown is a sample case study to determine which country has the most coal consumption and how much it changes over time.

```
# Import tidyverse and ggthemes
library(tidyverse)
library(ggthemes)

# Import coal consumption data file.
coal <- read_csv("http://594442.youcanlearnit.net/coal.csv", skip = 2)
```

Tidying Data

The data is all bundled up and random, so we are going to group it in terms of countries and regions.

```
# Check the data for coal
head(coal)
```

```
## # A tibble: 6 x 31
##   X1      `1980` `1981` `1982` `1983` `1984` `1985` `1986` `1987` `1988`
##   <chr> <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>
## 1 Nort~ 16.45~ 16.98~ 16.47~ 17.12~ 18.42~ 18.81~ 18.52~ 19.43~ 20.40~
## 2 Berm~ 0      0      0      0      0      0      0      0      0
## 3 Cana~ 0.961~ 0.990~ 1.055~ 1.116~ 1.236~ 1.206~ 1.125~ 1.250~ 1.358~
## 4 Gree~ 0.000~ 0.000~ 0.000~ 0.000~ 0.000~ 0      0      0      0
## 5 Mexi~ 0.102~ 0.105~ 0.119~ 0.128~ 0.130~ 0.146~ 0.156~ 0.170~ 0.159~
## 6 Sain~ 0      0      0      0      0      0      0      0      0
## # ... with 21 more variables: `1989` <chr>, `1990` <chr>, `1991` <chr>,
## #   `1992` <chr>, `1993` <chr>, `1994` <chr>, `1995` <chr>, `1996` <chr>,
```

```
## # `1997` <chr>, `1998` <chr>, `1999` <chr>, `2000` <chr>, `2001` <chr>,
## # `2002` <chr>, `2003` <chr>, `2004` <chr>, `2005` <chr>, `2006` <chr>,
## # `2007` <chr>, `2008` <chr>, `2009` <chr>
```

```
# Rename the First Column as it is undefined
colnames(coal)[1] <- 'region'
head(coal)
```

```
## # A tibble: 6 x 31
##   region `1980` `1981` `1982` `1983` `1984` `1985` `1986` `1987` `1988`
##   <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 North~ 16.45~ 16.98~ 16.47~ 17.12~ 18.42~ 18.81~ 18.52~ 19.43~ 20.40~
## 2 Bermu~ 0      0      0      0      0      0      0      0      0
## 3 Canada 0.961~ 0.990~ 1.055~ 1.116~ 1.236~ 1.206~ 1.125~ 1.250~ 1.358~
## 4 Green~ 0.000~ 0.000~ 0.000~ 0.000~ 0.000~ 0      0      0      0
## 5 Mexico 0.102~ 0.105~ 0.119~ 0.128~ 0.130~ 0.146~ 0.156~ 0.170~ 0.159~
## 6 Saint~ 0      0      0      0      0      0      0      0      0
## # ... with 21 more variables: `1989` <chr>, `1990` <chr>, `1991` <chr>,
## # `1992` <chr>, `1993` <chr>, `1994` <chr>, `1995` <chr>, `1996` <chr>,
## # `1997` <chr>, `1998` <chr>, `1999` <chr>, `2000` <chr>, `2001` <chr>,
## # `2002` <chr>, `2003` <chr>, `2004` <chr>, `2005` <chr>, `2006` <chr>,
## # `2007` <chr>, `2008` <chr>, `2009` <chr>
```

```
# Check the summary of the coal data.
glimpse(coal)
```

```
## Observations: 232
## Variables: 31
## $ region <chr> "North America", "Bermuda", "Canada", "Greenland", "Mex...
## $ `1980` <chr> "16.45179", "0", "0.96156", "0.00005", "0.10239", "0", ...
## $ `1981` <chr> "16.98772", "0", "0.99047", "0.00005", "0.10562", "0", ...
## $ `1982` <chr> "16.47546", "0", "1.05584", "0.00003", "0.11967", "0", ...
## $ `1983` <chr> "17.12407", "0", "1.11653", "0.00003", "0.12869", "0", ...
## $ `1984` <chr> "18.4267", "0", "1.23682", "0.00003", "0.13071", "0", ...
## $ `1985` <chr> "18.81819", "0", "1.20679", "0", "0.14646", "0", "17.46...
## $ `1986` <chr> "18.52559", "0", "1.12583", "0", "0.15609", "0", "17.24...
## $ `1987` <chr> "19.43781", "0", "1.25072", "0", "0.17001", "0", "18.01...
## $ `1988` <chr> "20.40363", "0", "1.35809", "0", "0.15967", "0", "18.88...
## $ `1989` <chr> "20.62571", "0", "1.35196", "0", "0.17359", "0", "19.10...
## $ `1990` <chr> "20.5602", "0", "1.21338", "0", "0.1694", "0", "19.1774...
## $ `1991` <chr> "20.4251", "0", "1.26457", "0", "0.15916", "0", "19.001...
## $ `1992` <chr> "20.64672", "0", "1.32379", "0", "0.16584", "0", "19.15...
## $ `1993` <chr> "21.28219", "0", "1.22875", "0", "0.19118", "0", "19.86...
## $ `1994` <chr> "21.39631", "0", "1.24492", "0", "0.1836", "0", "19.967...
## $ `1995` <chr> "21.64225", "0", "1.28479", "0", "0.20768", "0", "20.14...
## $ `1996` <chr> "22.57572", "0", "1.30032", "0", "0.25067", "0", "21.02...
## $ `1997` <chr> "23.20491", "0", "1.44933", "0", "0.26373", "0", "21.49...
## $ `1998` <chr> "23.5002", "0", "1.50985", "0", "0.26753", "0", "21.722...
## $ `1999` <chr> "23.4747", "0", "1.505", "0", "0.28947", "0", "21.68023...
## $ `2000` <chr> "24.55583", "0", "1.61651", "0", "0.29444", "0", "22.64...
## $ `2001` <chr> "23.62705", "0", "1.35444", "0", "0.32908", "0", "21.94...
## $ `2002` <chr> "23.69876", "0", "1.36876", "0", "0.36525", "0", "21.96...
## $ `2003` <chr> "24.17788", "0", "1.38766", "0", "0.41878", "0", "22.37...
## $ `2004` <chr> "24.36024", "0", "1.43684", "0", "0.31944", "0", "22.60...
## $ `2005` <chr> "24.6876", "0", "1.44948", "0", "0.39739", "0", "22.840...
## $ `2006` <chr> "24.32174", "0", "1.42135", "0", "0.39244", "0", "22.50...
```

```
## $ `2007` <chr> "24.54746", "0", "1.38369", "0", "0.38911", "0", "22.77...
## $ `2008` <chr> "24.11993", "0", "1.37388", "0", "0.32008", "0", "22.42...
## $ `2009` <chr> "21.14803", "0", "1.14314", "0", "0.3365", "0", "19.668...
```

```
# Use gather() to simplify the summary
coal_long <- gather(coal, 'year', 'coal_consumption', -region)
glimpse(coal_long)
```

```
## Observations: 6,960
## Variables: 3
## $ region      <chr> "North America", "Bermuda", "Canada", "Greenl...
## $ year        <chr> "1980", "1980", "1980", "1980", "1980", "1980...
## $ coal_consumption <chr> "16.45179", "0", "0.96156", "0.00005", "0.102...
```

```
# Convert year to date and coal value to numeric
coal_long$year <- as.integer(coal_long$year)
coal_long$coal_consumption <- as.numeric(coal_long$coal_consumption)
summary(coal_long)
```

```
##      region      year      coal_consumption
## Length:6960      Min.   :1980      Min.   : -0.0002
## Class :character  1st Qu.:1987      1st Qu.:  0.0000
## Mode  :character  Median :1994      Median :  0.0002
##                               Mean  :1994      Mean   :  1.3256
##                               3rd Qu.:2002      3rd Qu.:  0.0773
##                               Max.   :2009      Max.   :138.8298
##                               NA's   :517
```

```
# Locate regions and create columns of regions
head(unique(coal_long$region), n = 10)
```

```
## [1] "North America"      "Bermuda"
## [3] "Canada"             "Greenland"
## [5] "Mexico"             "Saint Pierre and Miquelon"
## [7] "United States"      "Central & South America"
## [9] "Antarctica"         "Antigua and Barbuda"
```

```
noncountries <- c("North America", "Central & South America", "Antarctica",
                  "Europe", "Eurasia", "Middle East", "Africa", "Asia & Oceania", "World")
```

```
# Match and filter data with noncountries and countries
matches <- which(!is.na(match(coal_long$region, noncountries)))

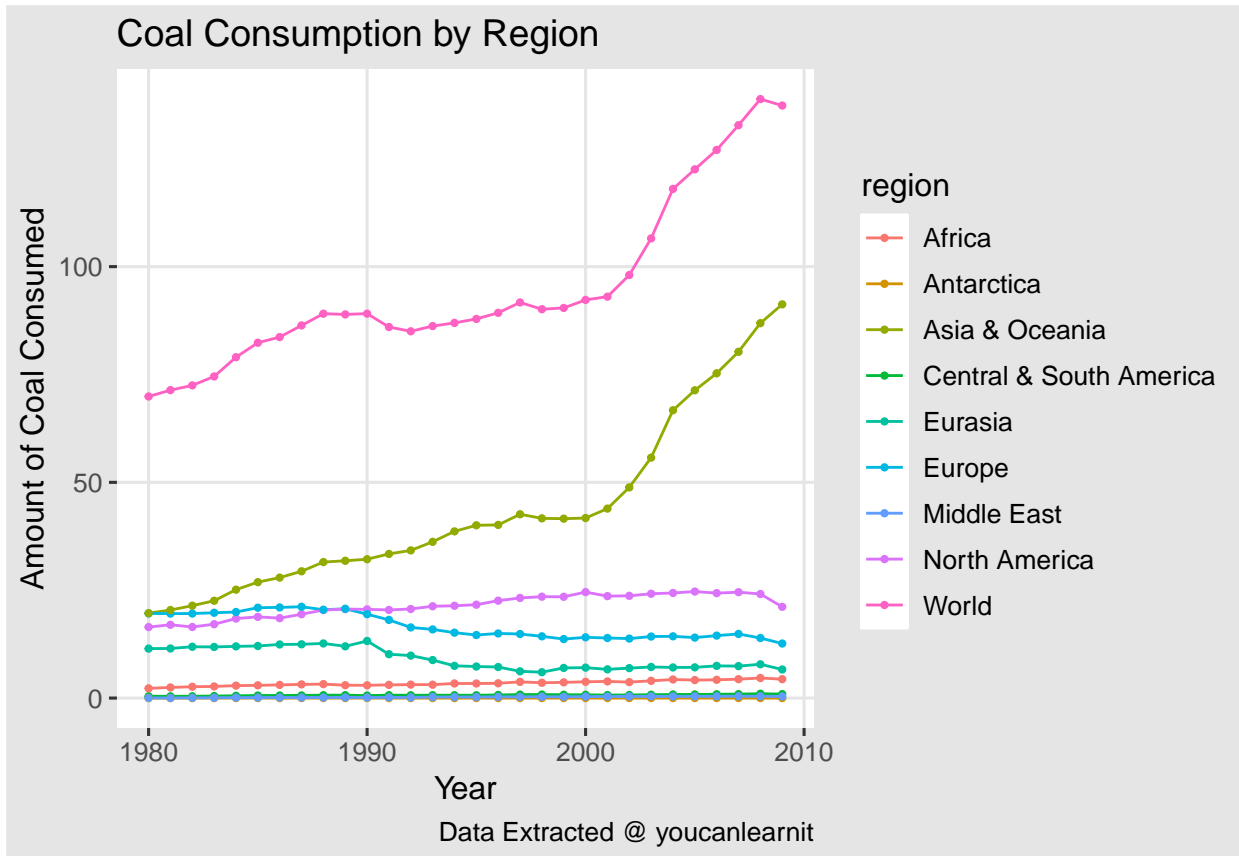
coal_country <- coal_long[-matches,]
coal_region <- coal_long[matches,]
```

Data Visualization

We are going to visualize which region has the most coal consumption. From what we can see, Asia and Oceania regions produces the most consumption of coal out of other regions. Moreover, the consumption begins to rise drastically in the year 2000.

```
ggplot(coal_region, aes(x = year, y = coal_consumption)) +
  geom_line(aes(color = region)) +
  geom_point(aes(color = region), size = 0.8) +
```

```
labs(title = "Coal Consumption by Region",
      caption = "Data Extracted @ youcanlearnit",
      x = "Year",
      y = "Amount of Coal Consumed") +
theme_igray()
```



Finding a Reason

Because Asia and Oceania is the major contributor of the global coal consumption, we are going to determine which country is causing the dramatic increase in coal consumption. Let's assume that we are looking for a major contributor in the year 2005.

```
# Find Countries that are causing a major coal consumption.
coal_country %>%
  filter(year %in% c(2005)) %>%
  arrange(desc(coal_consumption))
```

```
## # A tibble: 223 x 3
##   region      year coal_consumption
##   <chr>      <int>         <dbl>
## 1 China      2005          48.3
## 2 United States 2005          22.8
## 3 India       2005           8.61
## 4 Japan       2005           4.60
## 5 Russia      2005           4.29
```

```
## 6 South Africa 2005 3.81
## 7 Germany 2005 3.33
## 8 Australia 2005 2.31
## 9 Poland 2005 2.19
## 10 Korea, South 2005 2.07
## # ... with 213 more rows
```

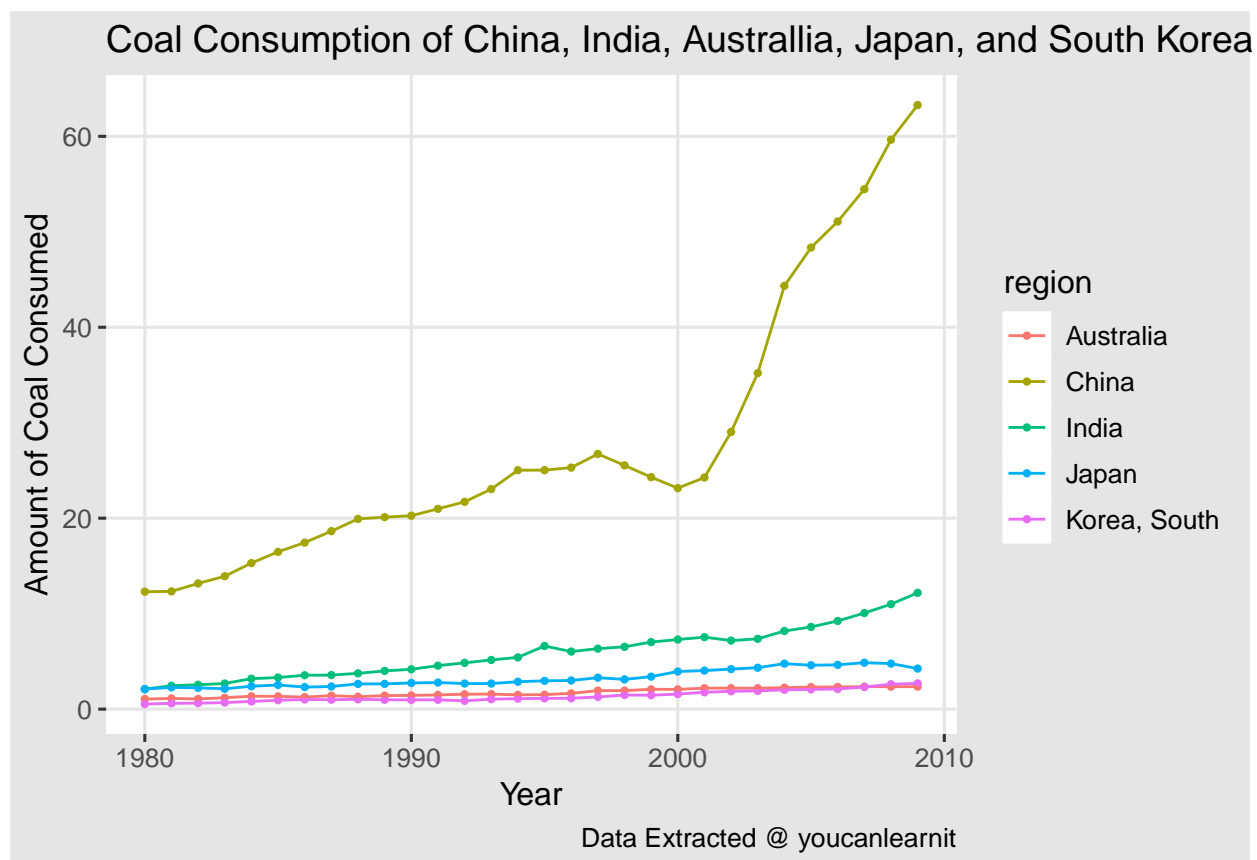
In this top 10 list, China and United States, by a landside, are the two major producers of the coal consumption. However, we are looking for Asia and Oceania, so we will filter the data based on this region.

```
coal_country %>%
  filter(region %in% c("China","India","Japan","Australia","Korea, South"))
```

```
## # A tibble: 150 x 3
##   region      year coal_consumption
##   <chr>      <int>         <dbl>
## 1 Australia  1980          1.05
## 2 China      1980         12.3
## 3 India      1980          2.09
## 4 Japan      1980          2.10
## 5 Korea, South 1980          0.533
## 6 Australia  1981          1.12
## 7 China      1981         12.3
## 8 India      1981          2.45
## 9 Japan      1981          2.28
## 10 Korea, South 1981          0.603
## # ... with 140 more rows
```

Visualizing Countries in Asia and Oceania

Now, we are going to create a graph based on five major coal consuming countries.

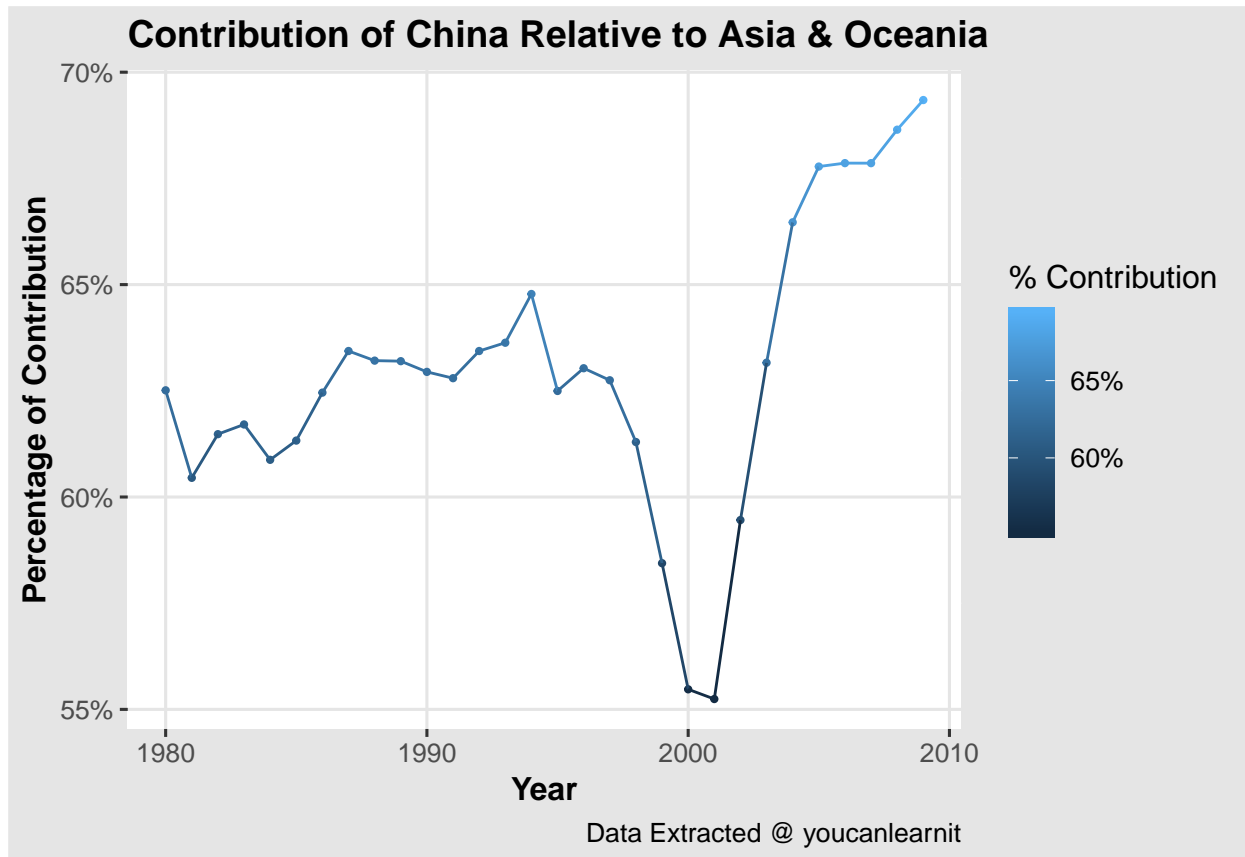


Ratio between China and Asia & Oceania Region

Compared to the overall Asia and Oceania Coal Consumption, how much ratio China holds?

```
# Create functions speicfically for China and Asia & Oceania
china_coal <- coal_country %>% filter(region == "China")
as_oc_coal <- coal_region %>% filter(region == "Asia & Oceania")
chin_coal_per <- china_coal %>% mutate(chin_perc = coal_consumption / as_oc_coal$coal_consumption)

# Plot a graph
ggplot(chin_coal_per, aes(x = year, y = chin_perc, color = chin_perc)) +
  geom_point(size = 0.8) +
  geom_line() +
  theme_igray() +
  labs(title = "Contribution of China Relative to Asia & Oceania",
       caption = "Data Extracted @ youcanlearnit",
       x = "Year",
       y = "Percentage of Contribution") +
  theme(plot.title = element_text(size = 14, face = "bold"),
        axis.title.x = element_text(size = 12, face = "bold"),
        axis.title.y = element_text(size = 12, face = "bold")) +
  scale_color_gradient(name = "% Contribution", labels = percent) +
  scale_y_continuous(labels = scales::percent)
```



It shows that the overall ratio between China and Asia and Oceania Region is at least 55% throughout the year.

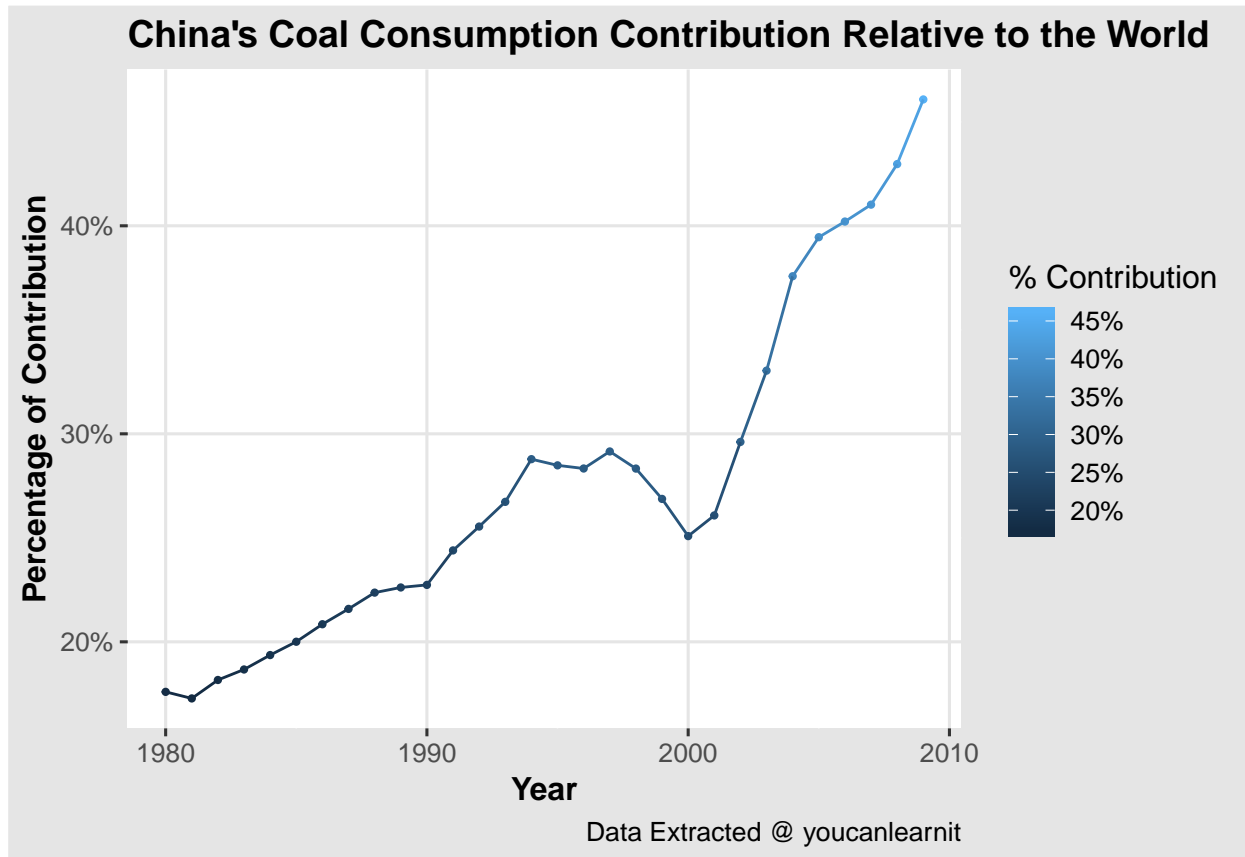
China vs the World

It seems that China holds the majority of the coal consumption in Asia and Oceania Region. Then, how does it compare to the global average?

```
# Create functions specifically for China and Asia & Oceania
china_coal <- coal_country %>% filter(region == "China")
world_coal <- coal_region %>% filter(region == "World")
chin_coal_per <- china_coal %>% mutate(chin_perc = coal_consumption / world_coal$coal_consumption)

# Plot a graph
ggplot(chin_coal_per, aes(x = year, y = chin_perc, color = chin_perc)) +
  geom_point(size = 0.8) +
  geom_line() +
  theme_igray() +
  labs(title = "China's Coal Consumption Contribution Relative to the World",
       caption = "Data Extracted @ youcanlearnit",
       x = "Year",
       y = "Percentage of Contribution"
  ) +
  theme(plot.title = element_text(size = 14, face = "bold"),
        axis.title.x = element_text(size = 12, face = "bold"),
        axis.title.y = element_text(size = 12, face = "bold")) +
```

```
scale_color_gradient(name = "% Contribution", labels = percent) +
scale_y_continuous(labels = scales::percent)
```



It seems that China has steadily increased its coal consumption since 1980. While there was a slight stagnancy around year 1996-2000, its coal consumption has drastically risen since then.

Conclusion

While the sample is part of the overall Coal Consumption data, the result does show that China contributes significantly in Coal Consumption compared to other countries or regions. It is an expected result because it has an overwhelming population and its number of citizens continues to rise significantly, contributing to a drastic increase in the coal consumption starting the year 2000.