

Untitled

Mike Marin

2019-09-17

Gapminder Data Summary

Data Description

The gapminder data contains information recorded for a number of countries

Variables recorded include: **Country, Continent, Year, Life Expectancy, Population, GDP Per Capita**

This is another slide added, without a title :)

I wanted to make sure I was able to add a slide, without giving it a title

Some Univariate Summaries

Univariate Summaries

It is of interest to take a look at each of the variables, and provide a quick summary, both numerically as well as graphically

Summary Statistics

##	country	continent	year	lifeExp
##	Afghanistan: 12	Africa :624	Min. :1952	Min. :23.60
##	Albania : 12	Americas:300	1st Qu.:1966	1st Qu.:48.20
##	Algeria : 12	Asia :396	Median :1980	Median :60.71
##	Angola : 12	Europe :360	Mean :1980	Mean :59.47
##	Argentina : 12	Oceania : 24	3rd Qu.:1993	3rd Qu.:70.85
##	Australia : 12		Max. :2007	Max. :82.60
##	(Other) :1632			
##	pop	gdpPercap		
##	Min. :6.001e+04	Min. : 241.2		
##	1st Qu.:2.794e+06	1st Qu.: 1202.1		
##	Median :7.024e+06	Median : 3531.8		
##	Mean :2.960e+07	Mean : 7215.3		
##	3rd Qu.:1.959e+07	3rd Qu.: 9325.5		
##	Max. :1.319e+09	Max. :113523.1		
##				

Countries and Continents

Which countries are represented in the data?

- ▶ In total there are 142 countries represented. They are not all listed here for the sake of space.
- ▶ There are a total of 5 geopolitical continents recorded:
 - ▶ Africa
 - ▶ Asia
 - ▶ Europe
 - ▶ Americas (North America and South America combined)
 - ▶ Oceania (Australia, New Zealand, and surrounding island nations)
 - ▶ Antarctica is not included in the data

Years Recorded

The data was collected for a number of years. The years the data was collected for were:

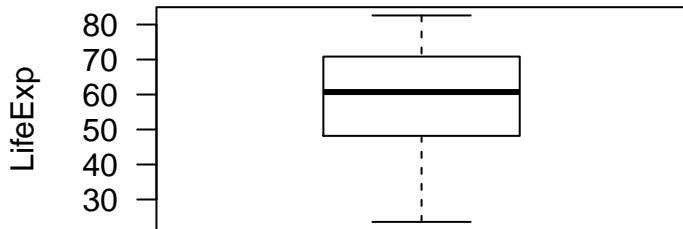
```
## [1] "1952" "1957" "1962" "1967" "1972" "1977" "1982"
```

```
## [1] "1987" "1992" "1997" "2002" "2007"
```

* Just adding some smaller font here for myself...

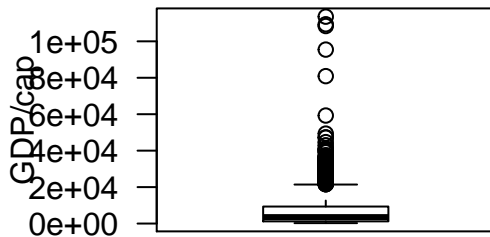
Life Expectancy

- Making a figure smaller (should probably center it too)



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	23.60	48.20	60.71	59.47	70.85	82.60

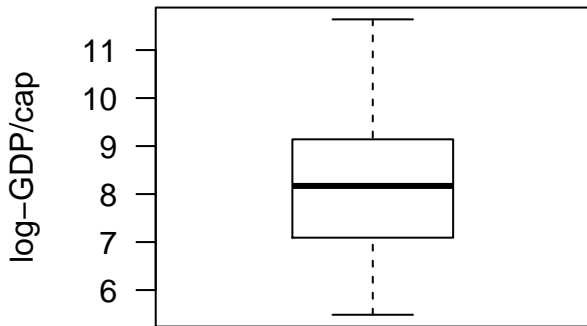
GDP Per Capita



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	241.2	1202.1	3531.8	7215.3	9325.5	113523.1

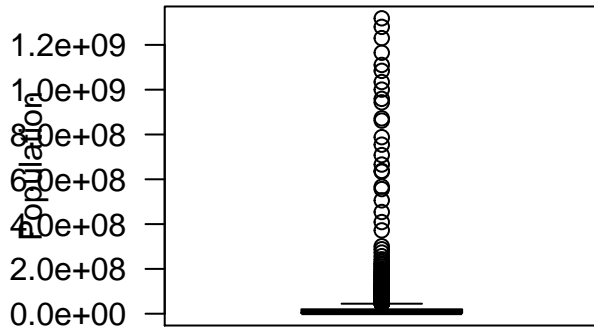
log-GDP (because it's a Rate)

A quick note that since GDP/cap is a **rate**, it may be better to be exploring it on the log(ln)-scale



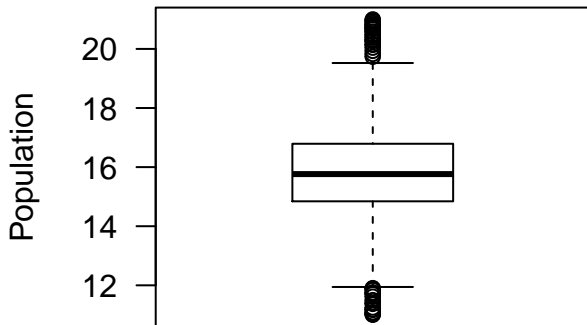
Population

Here are words...



log-Population

Again, it may be better to examine this on the log-scale, as populations tend to grow exponentially



Some Bivariate Summaries

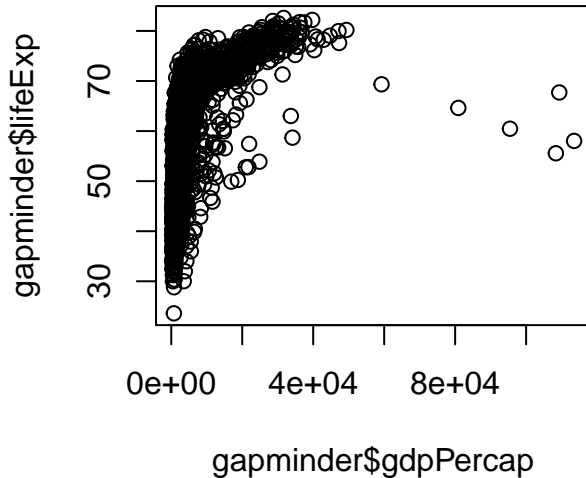
Bivariate Summaries

We may be interested in examining relationships between some of the variables. Following are a few of interest.

Note, we are ignoring the fact that the same countries have measurements taken for multiple years, for the time being. . .

Life Expectancy and GDP

It is reasonable to hypothesize that these would be related. Following is a visual examination of their relationship



Pearson's correlation

While the relationship clearly isn't linear, let's calculate this anyway...

```
cor(gapminder$gdpPercap, gapminder$lifeExp)
```

```
## [1] 0.5837062
```

Spearman's Correlation

Let's calculate Spearman's correlation to address the non-linearity, as it does appear to be a monotonic association:

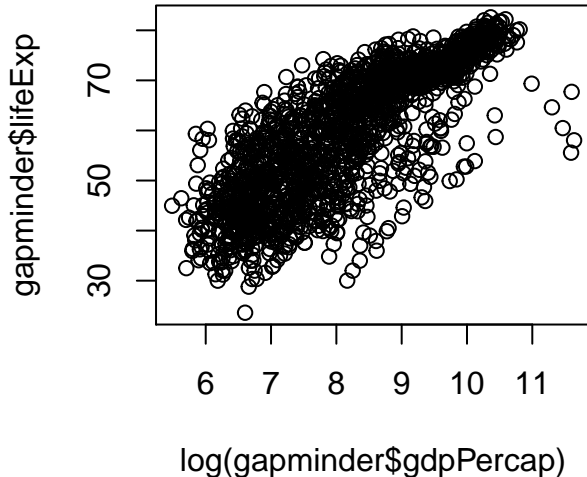
```
cor(gapminder$gdpPercap, gapminder$lifeExp, method="spearman")
```

```
## [1] 0.8264712
```

This makes more sense, but still, we can try other things...

Using log-GDP

Let's look at the same plot, but this time using the log-GDP-per-capita, as it makes sense to examine it on this scale



That Looked Like a Fix!

That looks like it would be easier to model. Let's also take a quick look at Pearson's correlation when we use log-GDP...

```
cor(log(gapminder$gdpPercap), gapminder$lifeExp, method="pearson")
```

```
## [1] 0.8076179
```

The End?

Yes, The End!

This will be the ending here

- ▶ There wasn't much focus on the content of presentation
- ▶ I wasn't consistent with using "echo=F" or not
 - ▶ But that's because I wanted to focus on learning more about how it formats slides, etc
- ▶ Some basic things were tried
- ▶ I don't really care about my grade, I just want to learn about the course material