STAT545 Asiggnemnt 1

Elijah Willie

September 18, 2018

Below are my exploratory analysis of the gapminder data

Load in the required libaries and the data

```
library(gapminder)
suppressMessages(library("tidyverse"))
```

Get some summary statistics on the data

```
summary(gapminder)
```

```
##         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
##
```

Add natural log of gdpPercap to our dataset

```
(gapminder.new = mutate(gapminder, lgdpPercap = log(gdpPercap)))
```

```
## # A tibble: 1,704 x 7
##    country     continent  year lifeExp      pop gdpPercap lgdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>      <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.       6.66
##  2 Afghanistan Asia       1957    30.3  9240934      821.       6.71
##  3 Afghanistan Asia       1962    32.0 10267083      853.       6.75
##  4 Afghanistan Asia       1967    34.0 11537966      836.       6.73
##  5 Afghanistan Asia       1972    36.1 13079460      740.       6.61
##  6 Afghanistan Asia       1977    38.4 14880372      786.       6.67
##  7 Afghanistan Asia       1982    39.9 12881816      978.       6.89
##  8 Afghanistan Asia       1987    40.8 13867957      852.       6.75
##  9 Afghanistan Asia       1992    41.7 16317921      649.       6.48
```

```
## 10 Afghanistan Asia       1997    41.8 22227415      635.       6.45
## # ... with 1,694 more rows
```

```
select(gapminder.new, contains("gdpPercap"))
```

```
## # A tibble: 1,704 x 2
##    gdpPercap lgdpPercap
##        <dbl>      <dbl>
##  1      779.       6.66
##  2      821.       6.71
##  3      853.       6.75
##  4      836.       6.73
##  5      740.       6.61
##  6      786.       6.67
##  7      978.       6.89
##  8      852.       6.75
##  9      649.       6.48
## 10      635.       6.45
## # ... with 1,694 more rows
```

Create a new dataset from year 1952 in Africa, Europe and Asia

```
(gm1952 = filter(gapminder.new, year==1952,
                 continent %in% c("Africa", "Europe", "Asia")))
```

```
## # A tibble: 115 x 7
##    country            continent  year lifeExp     pop gdpPercap lgdpPercap
##    <fct>              <fct>     <int>   <dbl>   <int>     <dbl>      <dbl>
##  1 Afghanistan        Asia       1952    28.8  8.43e6      779.       6.66
##  2 Albania            Europe     1952    55.2  1.28e6     1601.       7.38
##  3 Algeria            Africa     1952    43.1  9.28e6     2449.       7.80
##  4 Angola             Africa     1952    30.0  4.23e6     3521.       8.17
##  5 Austria            Europe     1952    66.8  6.93e6     6137.       8.72
##  6 Bahrain            Asia       1952    50.9  1.20e5     9867.       9.20
##  7 Bangladesh         Asia       1952    37.5  4.69e7      684.       6.53
##  8 Belgium            Europe     1952    68    8.73e6     8343.       9.03
##  9 Benin              Africa     1952    38.2  1.74e6     1063.       6.97
## 10 Bosnia and Herzeg~ Europe     1952    53.8  2.79e6      974.       6.88
## # ... with 105 more rows
```

Create a matching dataset for 2007

```
(gm2007 = filter(gapminder.new, year==2007,
                 continent %in% c("Africa", "Europe", "Asia")))
```

```
## # A tibble: 115 x 7
##    country           continent  year lifeExp      pop gdpPercap lgdpPercap
##    <fct>             <fct>     <int>   <dbl>    <int>     <dbl>      <dbl>
##  1 Afghanistan       Asia       2007    43.8   3.19e7      975.       6.88
##  2 Albania           Europe     2007    76.4   3.60e6     5937.       8.69
##  3 Algeria           Africa     2007    72.3   3.33e7     6223.       8.74
##  4 Angola            Africa     2007    42.7   1.24e7     4797.       8.48
##  5 Austria           Europe     2007    79.8   8.20e6    36126.      10.5
```

```
##  6 Bahrain           Asia       2007    75.6   7.09e5    29796.       10.3
##  7 Bangladesh        Asia       2007    64.1   1.50e8     1391.        7.24
##  8 Belgium           Europe     2007    79.4   1.04e7    33693.       10.4
##  9 Benin             Africa     2007    56.7   8.08e6     1441.        7.27
## 10 Bosnia and Herze~ Europe     2007    74.9   4.55e6     7446.        8.92
## # ... with 105 more rows
```

Plot life expectency vs log gdp per capita for the 3 continents on 1 axis in 1952

```
ggplot(gm1952, aes(x=lifeExp, y=lgdpPercap)) +
  geom_point(aes(color=continent)) +
  geom_smooth(method="lm", se=F, aes(colour=continent))
```