DS5110 HW 0

Kylie Bemis

1/8/2019

Problem 1

First, load the `ggplot2` visualization package.

```
library(ggplot2)
```

Print the dataset:

```
diamonds
```

```
## # A tibble: 53,940 x 10
##    carat cut       color clarity depth table price     x     y     z
##    <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
##  1 0.23  Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
##  2 0.21  Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
##  3 0.23  Good      E     VS1      56.9    65   327  4.05  4.07  2.31
##  4 0.290 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
##  5 0.31  Good      J     SI2      63.3    58   335  4.34  4.35  2.75
##  6 0.24  Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
##  7 0.24  Very Good I     VVS1     62.3    57   336  3.95  3.98  2.47
##  8 0.26  Very Good H     SI1      61.9    55   337  4.07  4.11  2.53
##  9 0.22  Fair      E     VS2      65.1    61   337  3.87  3.78  2.49
## 10 0.23  Very Good H     VS1      59.4    61   338  4     4.05  2.39
## # ... with 53,930 more rows
```

We can use the `ggplot2` package to visualize the `diamonds` dataset, which is included in the package:

```
ggplot(diamonds, aes(x=carat, y=price, color=clarity)) +
  geom_point() +
  geom_smooth() +
  labs(title="Diamond Price vs Carat")
```
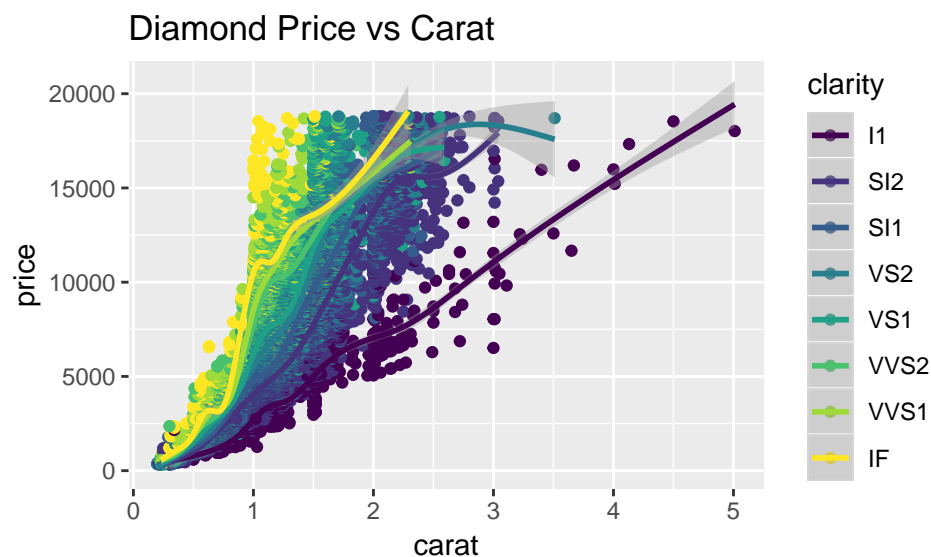
Problem 2

Now we load the `gapminder` package, which includes a dataset on life expectancy, GDP per capita, and population by country.

```
library(gapminder)
```

Print the dataset:

```
gapminder
```

```
## # A tibble: 1,704 x 6
##     country     continent  year lifeExp      pop gdpPercap
##     <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # ... with 1,694 more rows
```

Now we can use `ggplot2` to visualize the data:

```
ggplot(gapminder, aes(x=year, y=lifeExp, color=continent)) +
  geom_point() +
  geom_smooth() +
  labs(title="Life expectancy vs Year")
```