R Notebook

Code

#1

Hide

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(RColorBrewer)
```

Hide

```
forbes_billionaires_geo<- read.csv(file.choose())
```

Hide

```
forbes_billionaires_geo_sep<-forbes_billionaires_geo %>%separate(Education, c('UnderGrad','Grad'),sep = ";")
```

```
Expected 2 pieces. Additional pieces discarded in 66 rows [46, 92, 96, 158, 178, 270, 282, 300, 318, 338, 379, 408, 449, 480, 497, 525, 585, 643, 648, 649, ...].Expected 2 pieces. Missing pieces filled with `NA` in 2290 rows [1, 2, 3, 4, 5, 11, 12, 13, 15, 16, 17, 18, 22, 24, 26, 28, 29, 30, 31, 32, ...].
```

Hide

```
forbes_billionaires_geo_sep<-forbes_billionaires_geo_sep %>% separate(UnderGrad, c('UnderGradDegree','UnderGradUniversity'),sep = ",")
```

```
Expected 2 pieces. Additional pieces discarded in 53 rows [29, 100, 119, 146, 304, 309, 322, 326, 449, 490, 511, 525, 585, 618, 640, 677, 683, 813, 847, 1070, ...].Expected 2 pieces. Missing pieces filled with `NA` in 1370 rows [11, 12, 13, 24, 28, 34, 35, 37, 38, 40, 41, 42, 44, 47, 57, 58, 61, 62, 63, 70, ...].
```

Hide

```
forbes_billionaires_geo_sep<-forbes_billionaires_geo_sep %>% separate(Grad, c('GradDegree','GradUniversity'),sep = ",")
```

```
Expected 2 pieces. Additional pieces discarded in 40 rows [7, 9, 68, 95, 96, 100, 183, 351, 442, 449, 486, 490, 525, 533, 563, 569, 585, 597, 704, 723, ...].Expected 2 pieces. Missing pieces filled with `NA` in 2 rows [1145, 2615].
```

Hide

```
forbes_billionaires_geo_sep
```

Name <chr>	NetWo… <dbl>	Country <chr>	Source <chr>
Jeff Bezos	177.0	United States	Amazon
Elon Musk	151.0	United States	Tesla, SpaceX
Bernard Arnault & family	150.0	France	LVMH
Bill Gates	124.0	United States	Microsoft
Mark Zuckerberg	97.0	United States	Facebook
Warren Buffett	96.0	United States	Berkshire Hathaway
Larry Ellison	93.0	United States	software
Larry Page	91.5	United States	Google
Sergey Brin	89.0	United States	Google
Mukesh Ambani	84.5	India	diversified

1-10 of 2,755 rows | 1-6 of 16 columns Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
forbes_billionaires_geo_sep
```

Name <chr>	NetWo… <dbl>	Country <chr>	Source <chr>
Jeff Bezos	177.0	United States	Amazon
Elon Musk	151.0	United States	Tesla, SpaceX
Bernard Arnault & family	150.0	France	LVMH
Bill Gates	124.0	United States	Microsoft
Mark Zuckerberg	97.0	United States	Facebook
Warren Buffett	96.0	United States	Berkshire Hathaway
Larry Ellison	93.0	United States	software
Larry Page	91.5	United States	Google
Sergey Brin	89.0	United States	Google
Mukesh Ambani	84.5	India	diversified

1-10 of 2,755 rows | 1-6 of 16 columns Previous **1** 2 3 4 5 6 … 100 Next

#2 Is a scatterplot appropriate for any combination of variables in this dataset? I think a scatter plot would be appropriate for a plot of the age and net worth of each billionare. From this plot, we could see if older individuals tend to be richer.

If a scatterplot is appropriate, what are the predictor and response variables?
The predictor variable is age and the response variable is net worth. My hypothesis is that the billionares' net worth increases with age.

#3 After examining the data set answer the following questions in the white space of your notebook:
What relationship are you trying to show in the plot you are about to make?
What types of variables are you relating? What type of plot will do this?" In my plot I am trying to show a relationship between a billionares wealth and their marital status. Because of the large variance in the networth of the billionares I will be using a boxplot that show the different quartiles and its median.

#4

Hide

```
forbes_billionaires_geo_sep
```

Name <chr>	NetWo… <dbl>	Country <chr>	Source <chr>
Jeff Bezos	177.0	United States	Amazon
Elon Musk	151.0	United States	Tesla, SpaceX
Bernard Arnault & family	150.0	France	LVMH
Bill Gates	124.0	United States	Microsoft
Mark Zuckerberg	97.0	United States	Facebook
Warren Buffett	96.0	United States	Berkshire Hathaway
Larry Ellison	93.0	United States	software
Larry Page	91.5	United States	Google
Sergey Brin	89.0	United States	Google
Mukesh Ambani	84.5	India	diversified

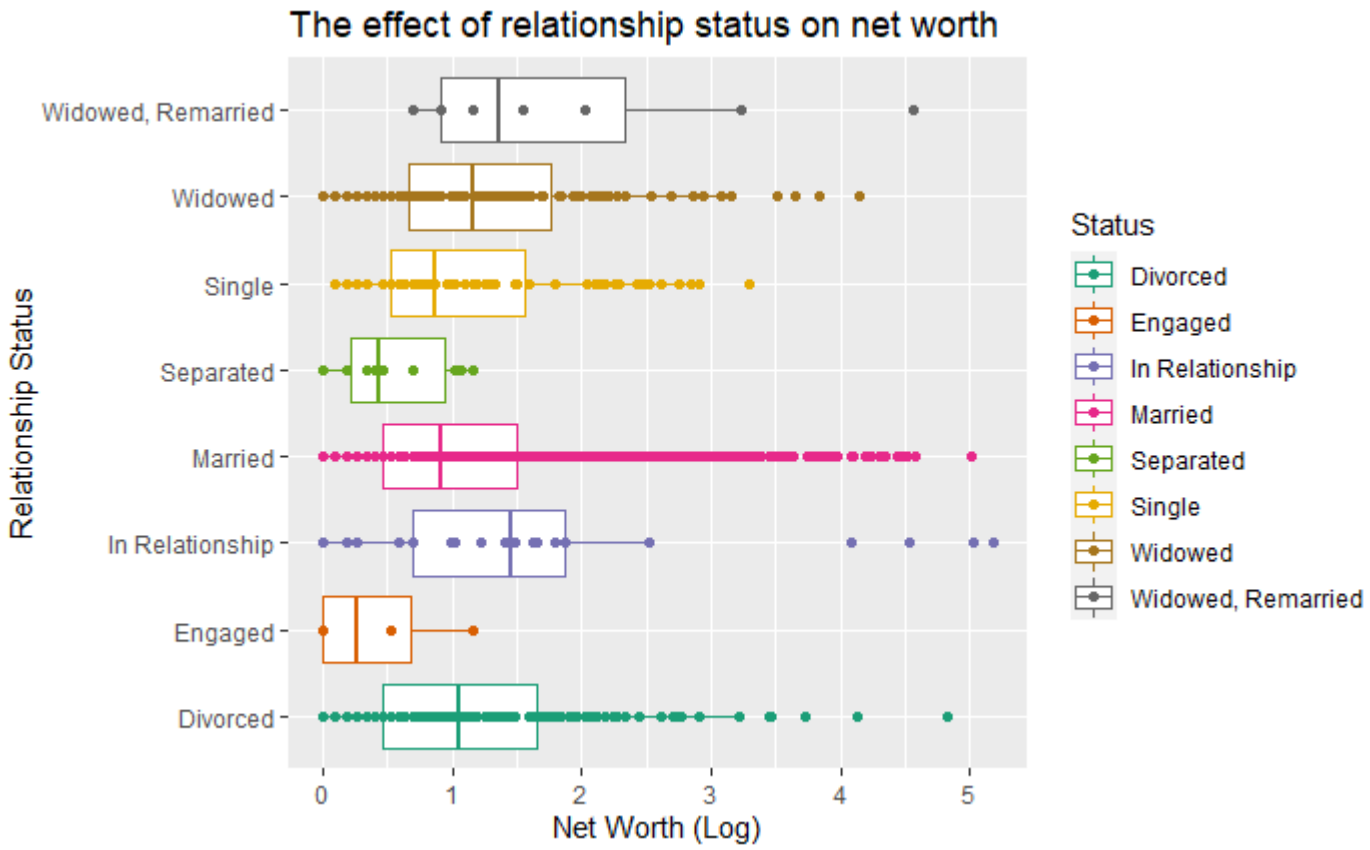
1-10 of 2,090 rows | 1-5 of 16 columns Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
forbes_billionaires_geo_sep=filter(forbes_billionaires_geo_sep,Status!="")
```

Hide

```
ggplot(forbes_billionaires_geo_sep,aes(x=log(NetWorth),y=Status,color=Status))+geom_boxplot()+ge
om_point()+scale_color_brewer(palette="Dark2")+labs(title="The effect of relationship status on
 net worth",x="Net Worth (Log)",y="Relationship Status")
```



The plot is showing a few key details about the data. Firstly, the group with the largest amount of data in it is the married group. This means that the relationship group that is most common for bilionares is married. This graph also shows that people in a relationship tend to be the most affluent, i.e. have the highest median net worth.