R Notebook

Principles of Data Visualization and Introduction to ggplot2

I have provided you with data about the 5,000 fastest growing companies in the US, as compiled by Inc. magazine. lets read this in:

```
inc <- read.csv("https://raw.githubusercontent.com/charleyferrari/CUNY_DATA_608/master/module1/Data/inc5
```

And lets preview this data:

```
head(inc)
```

```
##   Rank                         Name Growth_Rate   Revenue
## 1    1                         Fuhu      421.48 1.179e+08
## 2    2          FederalConference.com      248.31 4.960e+07
## 3    3                The HCI Group      245.45 2.550e+07
## 4    4                       Bridger      233.08 1.900e+09
## 5    5                        DataXu      213.37 8.700e+07
## 6    6   MileStone Community Builders      179.38 4.570e+07
##                         Industry Employees         City State
## 1 Consumer Products & Services       104   El Segundo    CA
## 2          Government Services        51     Dumfries    VA
## 3                      Health       132 Jacksonville    FL
## 4                      Energy        50      Addison    TX
## 5       Advertising & Marketing       220       Boston    MA
## 6                 Real Estate        63       Austin    TX
```

```
summary(inc)
```

```
##       Rank                         Name         Growth_Rate
##  Min.   :   1   (Add)ventures        :   1   Min.   :  0.340
##  1st Qu.:1252   @Properties          :   1   1st Qu.:  0.770
##  Median :2502   1-Stop Translation USA:   1   Median :  1.420
##  Mean   :2502   110 Consulting       :   1   Mean   :  4.612
##  3rd Qu.:3751   11thStreetCoffee.com :   1   3rd Qu.:  3.290
##  Max.   :5000   123 Exteriors        :   1   Max.   :421.480
##                 (Other)              :4995
##     Revenue                         Industry       Employees
##  Min.   :2.000e+06   IT Services             : 733   Min.   :    1.0
##  1st Qu.:5.100e+06   Business Products & Services: 482   1st Qu.:   25.0
##  Median :1.090e+07   Advertising & Marketing    : 471   Median :   53.0
##  Mean   :4.822e+07   Health                  : 355   Mean   :  232.7
##  3rd Qu.:2.860e+07   Software                : 342   3rd Qu.:  132.0
##  Max.   :1.010e+10   Financial Services      : 260   Max.   :66803.0
##                      (Other)                 :2358   NA's   :12
```

```
##            City             State
##  New York      : 160   CA     : 701
##  Chicago       :  90   TX     : 387
##  Austin        :  88   NY     : 311
##  Houston       :  76   VA     : 283
##  San Francisco :  75   FL     : 282
##  Atlanta       :  74   IL     : 273
##  (Other)       :4438   (Other):2764
```

Think a bit on what these summaries mean. Use the space below to add some more relevant non-visual exploratory information you think helps you understand this data:

```r
# Insert your code here, create more chunks as necessary

# Maximum Growth Rate
(Growth <- inc[which.max(inc$Growth_Rate),])
```

```
##   Rank Name Growth_Rate    Revenue                       Industry Employees
## 1    1 Fuhu      421.48 117900000 Consumer Products & Services       104
##          City State
## 1 El Segundo    CA
```

```r
# Maximum Revenue
(Revenue <- inc[which.max(inc$Revenue),])
```

```
##        Rank Name Growth_Rate  Revenue          Industry Employees
## 4788 4788  CDW         0.41 1.01e+10 Computer Hardware      6800
##             City State
## 4788 Vernon Hills    IL
```

```r
# Maximum Employees
(Employees <- inc[which.max(inc$Employees),])
```

```
##        Rank                        Name Growth_Rate    Revenue
## 2344 2345 Integrity staffing Solutions        1.55 278200000
##              Industry Employees       City State
## 2344 Human Resources      66803 Wilmington    DE
```