HW1

Dan Fanelli

February 2, 2017

Some Setup, Peek at the Core Data:

```r
library(knitr)
library(sqldf)
library(ggplot2)

data <- read.csv("inc5000_data.csv", header = TRUE)
kable(head(data))
```

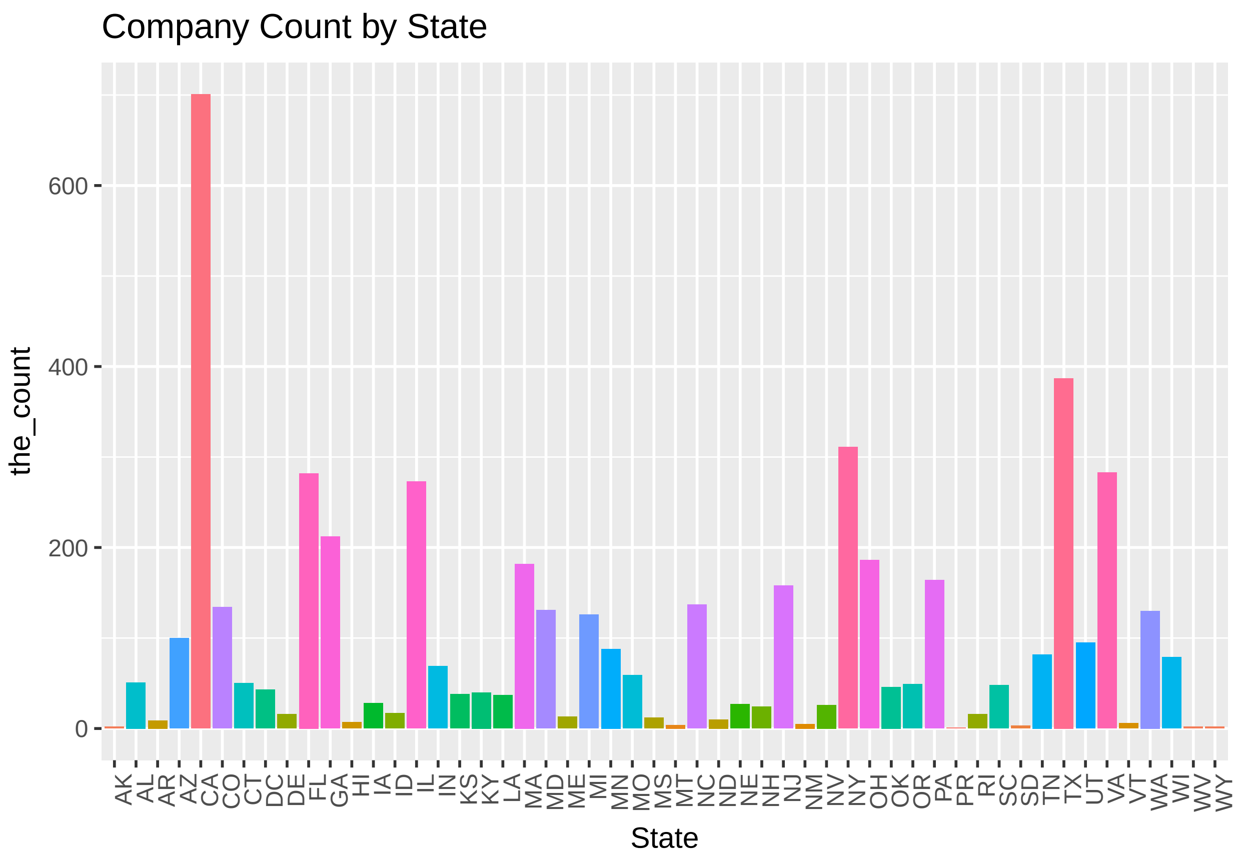
Rank	Name	Growth_Rate	Revenue	Industry	Employees	City
1	Fuhu	421.48	1.179e+08	Consumer Products & Services	104	El Se
2	FederalConference.com	248.31	4.960e+07	Government Services	51	Dum
3	The HCI Group	245.45	2.550e+07	Health	132	Jacks
4	Bridger	233.08	1.900e+09	Energy	50	Addi
5	DataXu	213.37	8.700e+07	Advertising & Marketing	220	Bost
6	MileStone Community Builders	179.38	4.570e+07	Real Estate	63	Aust

1) Companies by State

```r
state_count <- sqldf("select State, count(*) as the_count from data group by State order by the_count d
kable(head(state_count))
```

State	the_count
CA	701
TX	387
NY	311
VA	283
FL	282
IL	273

```r
ggplot(data=state_count, aes(x=State, y=the_count, fill=factor(the_count))) + geom_bar(stat="identity")
```

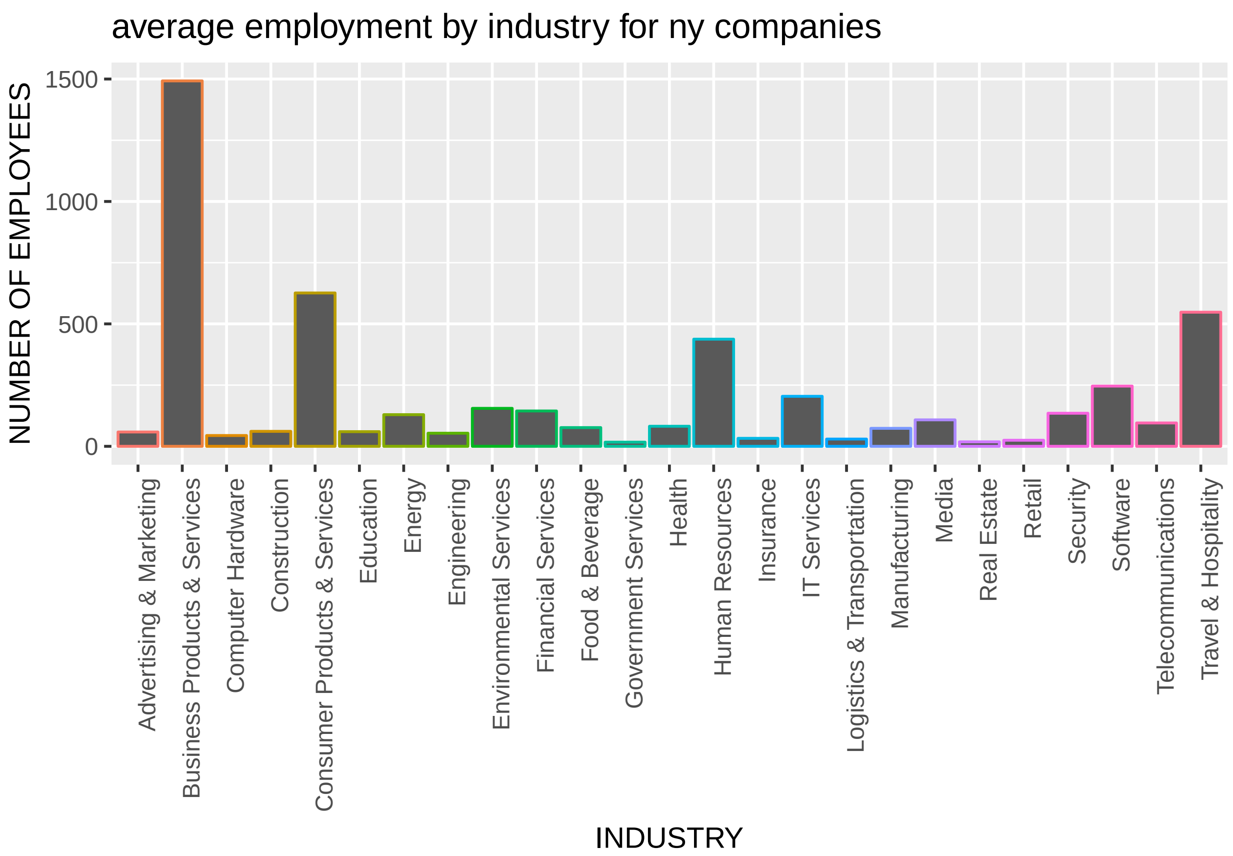


2) Average employment by industry for companies in state with the 3rd Most:

```
ny_data <- data[data$State == state_count$State[3],]
ny_data <- ny_data[complete.cases(ny_data),]
ny_data <- sqldf('select Industry, avg(Employees) as avg_employees, state from ny_data group by Industr
kable(head(ny_data))
```

Industry	avg_employees	State
Business Products & Services	1492.4615	NY
Consumer Products & Services	626.2941	NY
Travel & Hospitality	547.7143	NY
Human Resources	437.5455	NY
Software	245.9231	NY
IT Services	204.0930	NY

```
ggplot(ny_data, aes(x = factor(Industry), y = avg_employees)) + geom_bar(stat = "identity", aes(colour
```



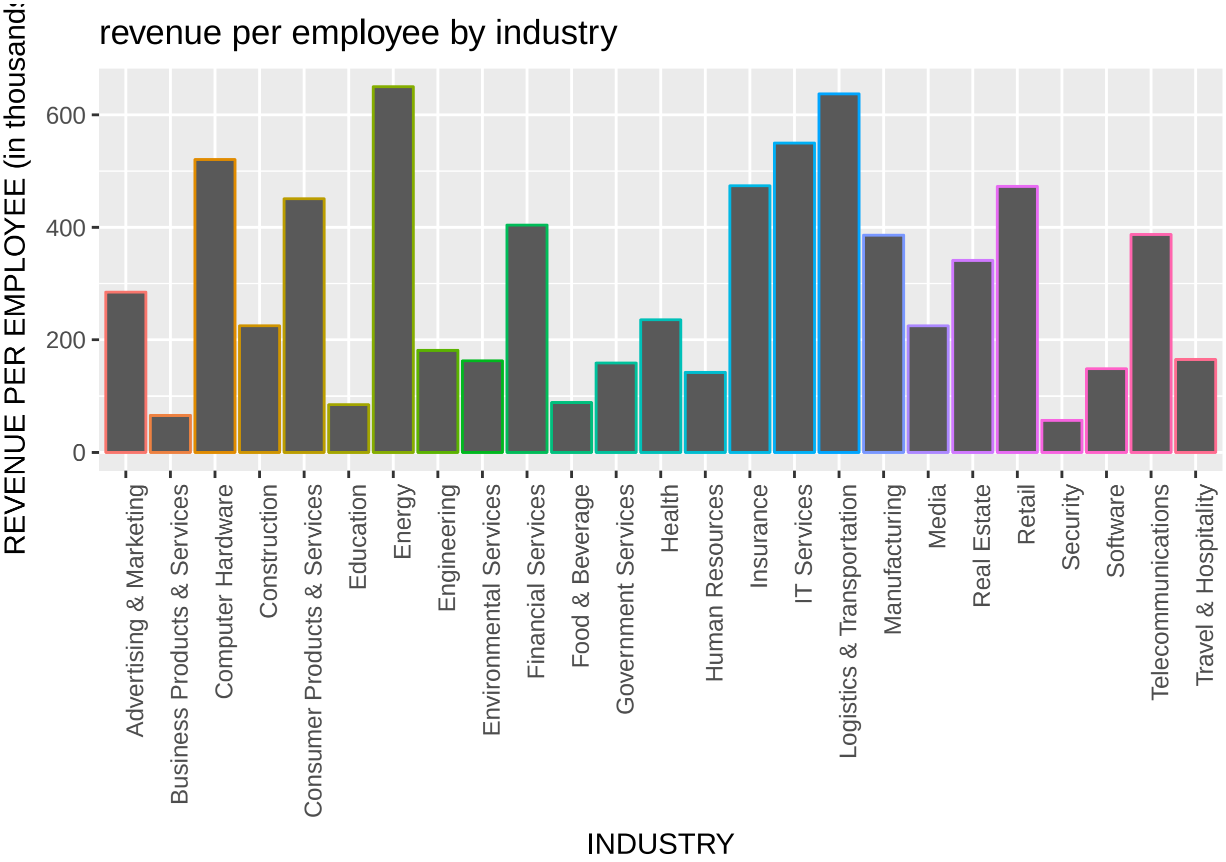
3) Which industries generate the most revenue per employee:

```
ny_data <- data[data$State == state_count$State[3],]
ny_data <- ny_data[complete.cases(ny_data),]

rev_data <- sqldf("select Industry, ((sum(Revenue)/sum(Employees))/1000) as revenue_per_employee from n
kable(head(rev_data))
```

Industry	revenue_per_employee
Energy	650.0000
Logistics & Transportation	637.2881
IT Services	549.9316
Computer Hardware	520.4545
Insurance	473.8462
Retail	472.6225

```
ggplot(rev_data, aes(x = factor(Industry), y = revenue_per_employee)) + geom_bar(stat = "identity", aes
```


revenue per employee by industry
REVENUE PER EMPLOYEE (in thousands
600
400
200
0
Advertising & Marketing
Business Products & Services
Computer Hardware
Construction
Consumer Products & Services
Education
Energy
Engineering
Environmental Services
Financial Services
Food & Beverage
Government Services
Health
Human Resources
Insurance
IT Services
Logistics & Transportation
Manufacturing
Media
Real Estate
Retail
Security
Software
Telecommunications
Travel & Hospitality
INDUSTRY