Banu Boopalan: Module1 Homework

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/inc</pre>

And lets preview this data:

head(inc)

```
##
     Rank
                                    Name Growth_Rate
                                                        Revenue
## 1
                                               421.48 1.179e+08
        1
                                    Fuhu
## 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
                                  DataXu
                                               213.37 8.700e+07
## 5
## 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
                                          132 Jacksonville
                                                                FL
                            Health
## 4
                            Energy
                                           50
                                                    Addison
                                                                TX
## 5
                                          220
          Advertising & Marketing
                                                                MA
                                                     Boston
## 6
                       Real Estate
                                           63
                                                     Austin
                                                                TX
```

summary(inc)

```
##
         Rank
                        Name
                                         Growth_Rate
                                                              Revenue
                    Length:5001
##
   Min.
                1
                                                : 0.340
                                                           Min.
                                                                   :2.000e+06
                                                           1st Qu.:5.100e+06
   1st Qu.:1252
                    Class : character
                                        1st Qu.:
                                                  0.770
                    Mode :character
##
    Median:2502
                                        Median:
                                                   1.420
                                                           Median :1.090e+07
##
    Mean
           :2502
                                        Mean
                                                   4.612
                                                           Mean
                                                                   :4.822e+07
    3rd Qu.:3751
##
                                        3rd Qu.:
                                                   3.290
                                                           3rd Qu.:2.860e+07
           :5000
    Max.
                                                :421.480
                                                                   :1.010e+10
##
                                        Max.
                                                           Max.
##
##
      Industry
                          Employees
                                                City
                                                                   State
    Length:5001
                        Min.
                                     1.0
                                           Length:5001
                                                               Length:5001
##
    Class : character
                        1st Qu.:
                                    25.0
                                           Class : character
                                                               Class : character
                                    53.0
##
    Mode :character
                        Median:
                                           Mode :character
                                                               Mode : character
##
                                   232.7
                        Mean
##
                        3rd Qu.: 132.0
                                :66803.0
##
                        Max.
                        NA's
##
                               :12
```

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:

##Answer tried to understand by summarizing to understand he data.

```
#install.packages("kableExtra")
library(kableExtra)
## Warning: package 'kableExtra' was built under R version 4.1.3
library(dplyr)
## Attaching package: 'dplyr'
## The following object is masked from 'package:kableExtra':
##
##
      group_rows
## The following objects are masked from 'package:stats':
##
##
      filter, lag
## The following objects are masked from 'package:base':
##
      intersect, setdiff, setequal, union
#install.packages("ggthemes")
library(ggthemes)
## Warning: package 'ggthemes' was built under R version 4.1.3
str(inc)
## 'data.frame':
                   5001 obs. of 8 variables:
## $ Rank
               : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Name
              : chr "Fuhu" "FederalConference.com" "The HCI Group" "Bridger" ...
## $ Growth_Rate: num 421 248 245 233 213 ...
## $ Revenue
               : num 1.18e+08 4.96e+07 2.55e+07 1.90e+09 8.70e+07 ...
## $ Industry : chr
                      "Consumer Products & Services" "Government Services" "Health" "Energy" ...
## $ Employees : int 104 51 132 50 220 63 27 75 97 15 ...
## $ City
                : chr "El Segundo" "Dumfries" "Jacksonville" "Addison" ...
                : chr "CA" "VA" "FL" "TX" ...
## $ State
library(tidyverse)
## Warning: package 'tidyverse' was built under R version 4.1.2
## -- Attaching packages ------ tidyverse 1.3.1 --
```

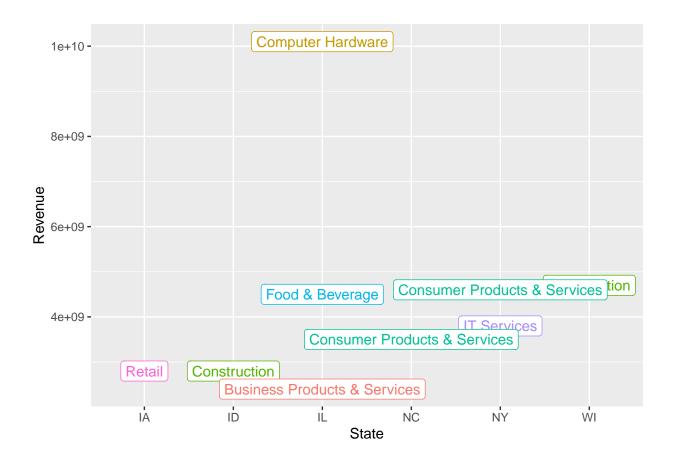
```
v purrr 0.3.4
## v ggplot2 3.3.5
## v tibble 3.1.5
                        v stringr 1.4.0
                        v forcats 0.5.1
## v tidyr
             1.1.4
## v readr
             2.1.1
## Warning: package 'readr' was built under R version 4.1.2
## Warning: package 'forcats' was built under R version 4.1.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()
                          masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()
                          masks stats::lag()
# Tufe theme. Quick graph showing industries with greater than 1000 employees along with revenue
ggplot(filter(inc,Employees > 1000 ), aes(x = Industry, y = Revenue)) +
  geom_point() +
  labs(x = "Industry", y = "Revenue") +
   theme_tufte() + coord_flip()
           Travel & Hospitality -
           Telecommunications -
                    Software -
                    Security -
                      Retail -
                  Real Estate -
                      Media -
               Manufacturing -
      Logistics & Transportation -
                  IT Services -
                   Insurance -
             Human Resources -
                     Health -
           Government Services -
             Food & Beverage -
             Financial Services -
         Environmental Services -
                 Engineering -
                     Energy -
   Consumer Products & Services -
                 Construction -
           Computer Hardware -
    Business Products & Services -
        Advertising & Marketing -
                                                                         7.5e+09
                                                                                        1.0e+10
                           0.0e+00
                                           2.5e+09
                                                          5.0e+09
                                                         Revenue
# Tufe theme. Quick graph showing industries with greater than 100 employees along with revenue
ggplot(filter(inc,Employees < 100 ), aes(x = Industry, y = Revenue)) +</pre>
  geom_point(fill = "indianred") +
  labs(x = "Industry", y = "Revenue") +
```

theme_tufte() + coord_flip()

```
Travel & Hospitality -
            Telecommunications -
                     Software -
                      Security -
                       Retail -
                   Real Estate -
                       Media -
                Manufacturing -
       Logistics & Transportation -
                   IT Services -
                     Insurance -
Industry
              Human Resources -
                       Health -
           Government Services -
              Food & Beverage -
              Financial Services -
         Environmental Services -
                   Engineering -
                       Energy -
                    Education -
    Consumer Products & Services -
                  Construction -
            Computer Hardware -
     Business Products & Services -
        Advertising & Marketing -
                                                                1.0e+09
                                                                                 1.5e+09
                                              5.0e+08
                             0.0e+00
                                                             Revenue
#Top 50 records by revenue
revenuetop50 = inc %>% arrange(desc(Revenue)) %>% head(50)
#revenuetop50
revenuetop50$Revenue1 = sapply(revenuetop50$Revenue, function(x) paste(round((x / 1e9), 1), "Billion")
revenuetop50 %>% kable() %>% kable_styling()
#print
sample <- revenuetop50 %>% arrange(desc(Revenue)) %>% head(10)
#describe unique value to find unique industries in the sample data
sample
##
      Rank
                                     Name Growth_Rate Revenue
      4788
## 1
                                      CDW
                                                   0.41 1.01e+10
## 2
      3853
                              ABC Supply
                                                   0.73 4.70e+09
## 3
      4936
                                                   0.36 4.60e+09
                                     Coty
## 4
      4997
                                Dot Foods
                                                   0.34 4.50e+09
## 5
      4716
                           Westcon Group
                                                   0.44 3.80e+09
## 6
      4246 American Tire Distributors
                                                   0.59 3.50e+09
## 7
      4052
                                 Kum & Go
                                                   0.65 2.80e+09
## 8
      4802
                           Boise Cascade
                                                   0.41 2.80e+09
## 9
      1397
                      EnvisionRxOptions
                                                   2.88 2.70e+09
## 10 2522
                                                   1.41 2.40e+09
                                DLA Piper
##
                             Industry Employees
                                                                                  Revenue1
                                                                City State
## 1
                   Computer Hardware
                                             6800
                                                       Vernon Hills
                                                                         IL 10.1 Billion
```

APS CDW	Rank	Name	Growth_Rate	Revenue	Industry	Employees
4997 Dot Foods	4788	CDW	0.41	1.010e+10	Computer Hardware	6800
4997 Dot. Foods 0.34 4.500e+09 Nool & Reverage 3919 4716 Westcon Group 0.44 3.800e+09 To Services 3304 4716 Westcon Group 0.45 3.800e+09 Consumer Products & Services 3341 4052 Kum & Go 0.65 2.800e+09 Consumer Products & Services 3341 4489 Retail 4889 Soise Cascade 0.41 2.800e+09 Construction 4470 44	3853	ABC Supply	0.73	4.700e+09	Construction	6549
4216 Mesteon Group	4936	Coty	0.36	4.600e+09	Consumer Products & Services	10000
4246 American Tire Distributors 0.59 3.500e+09 Consumer Products & Services 3341 4852 Boise Cascade 0.41 2.800e+09 Construction 4470 1397 EnvisionRxOptions 2.88 2.700e+09 Health 625 2522 DLA Piper 1.41 2.400e+09 Business Products & Services 4036 4629 Prine Therapeutics 0.47 2.000e+09 Health 2549 487 Bridger 233.08 1.900e+09 Health 2549 488 Bridger 233.08 1.900e+09 Energy 5.50 1843 Sun Coast Resources 2.08 1.900e+09 Energy 1.640 3844 Atlas Oil Company 0.74 1.900e+09 Energy 1.640 3844 Atlas Oil Company 0.74 1.900e+09 Ungisted & Transportation 374 4868 Global Brass and Copper Holdings 0.45 1.700e+09 Ungisted & Services 13200 4869 Global Brass and Copper Holdings 0.45 1.700e+09 Ungisted & Services 13200 2145 AdvancePierre Foods 1.73 1.500e+09 Business Products & Services 2280 2145 AdvancePierre Foods 1.73 1.500e+09 Pood & Beverage 4000 3600 Genco 0.82 1.500e+09 Dogisted & Transportation 10800 3601 Advanced Disposal 4.51 1.400e+09 Environmental Services 5347 3414 Restoration Hardware 0.91 1.200e+09 Environmental Services 3365 3661 Restoration Hardware 0.91 1.200e+09 Retail 2900 3734 Restoration Hardware 0.91 1.200e+09 Retail 2900 3734 Restoration Hardware 0.91 1.200e+09 Health 761 3734 KSS 0.78 1.000e+09 Health 761 3734 KSS 0.78 1.000e+09 Health 761 3735 LORD Corporation 0.44 8.615e+08 ImmixGroup 1.15 8.836e+08 IT Services 3369 3469 Ambit Energy 1.88 9.303e+08 Energy 492 3472 LORD Corporation 0.44 8.615e+08 University 492 3473 LORD Corporation 0.44 8.615e+08 University 492 3480 Coyote Logistics 2.70 7.864e+08 Energy 492 3491 Color English 4.90 4.90 4.90 4.90 4.90 4.90 3480 Coyote Logistics 2.70 7.864e+08 5.400 3480 Coyote Logistics 2.70 7.864e+08 5.400 5.	4997	Dot Foods	0.34	4.500e+09	Food & Beverage	3919
4802 Roise Cascade 0.41 2.800e+09 Construction 4470 1397 EnvisionRxOptions 2.88 2.700e+09 Health 625 2522 DLA Piper 1.41 2.400e+09 Business Products & Services 4036 4629 Prime Therapeutics 0.47 2.000e+09 Business Products & Services 4036 4629 Arrive Therapeutics 0.47 2.000e+09 Business Products & Services 4036 481 Bridger 233.08 1.900e+09 Energy 5.0 483 Sun Coast Resources 2.08 1.900e+09 Energy 1640 3844 Atlas Oil Company 0.74 1.900e+09 Logistics & Transportation 374 4861 Kirkland & Ellis 0.36 1.900e+09 Logistics & Services 1517 488 Sprouts Farmers Market 2.68 1.800e+09 Counter Products & Services 15200 488 Global Brass and Copper Holdings 0.45 1.700e+09 Manufacturing 1986 3463 Hogan Lovells 0.89 1.600e+09 Business Products & Services 2280 452 AvancePierre Foods 1.73 1.500e+09 Food & Beverage 4000 3550 Genco 0.82 1.500e+09 Logistics & Transportation 10800 3650 Genco 0.82 1.500e+09 Logistics & Transportation 10800 3650 Genco 0.81 1.400e+09 Engistics & Transportation 10800 3650 Genco 0.82 1.500e+09 Logistics & Transportation 2116 2496 Carabsoft Technology 1.42 1.400e+09 Engistics & Transportation 2116 2496 Carabsoft Technology 1.42 1.400e+09 Engistics & Transportation 2116 3734 KSS 0.78 1.000e+09 Retail 22000 3910 Diplomat Specialty Planmacy 1.99 1.100e+09 Retail 22000 3920 Ambit Energy 1.88 9.303e+08 Energy 492 3930 Ambit Energy 1.88 9.303e+08 Energy 492 3940 Admit Energy 1.88 9.303e+08 Energy 492 3941 Rotoroid 1.94 1.96 1.96 1.96 1.96 1.96 3941 Rotoroid 1.94 1.96 1.96 1.96 1.96 1.96 3941 Control 1.94 1.96 1.96 1.96 1.96 1.96 3942 Globady coon 0.49 9.109e+09 Retail 2.900 3943 Admit Retail 1.90 1.90 1.90	4716	Westcon Group	0.44	3.800e+09	IT Services	3000
4802 Boise Cascade	4246	American Tire Distributors	0.59	3.500e+09	Consumer Products & Services	3341
1397 EnvisiontROptions 2.88 2.700e+09 Health 625 2522 DLA Piper 1.44 2.400e+09 Business Products & Services 4036 4629 Prime Therapeutics 0.47 2.000e+09 Health 2549 4 Bridger 233.08 1.900e+09 Energy 50 3344 Atlas Oil Company 0.74 1.900e+09 Energy 1640 3344 Atlas Oil Company 0.74 1.900e+09 Energy 1640 3446 Kirkland & Ellis 0.36 1.900e+09 Business Products & Services 1517 4861 Kirkland & Ellis 0.36 1.900e+09 Business Products & Services 1527 4868 Global Brass and Copper Holdings 0.45 1.700e+09 Manufacturing 1986 4869 Global Brass and Copper Holdings 0.45 1.700e+09 Manufacturing 1986 4363 Hogan Lovells 0.88 1.600e+09 Consumer Products & Services 13200 4689 Global Brass and Copper Holdings 0.45 1.700e+09 Manufacturing 1986 4360 Genco 0.82 1.500e+09 Gode & Beverage 4000 4050 Genco 0.82 1.500e+09 Logistics & Transportation 10800 4050 Genco 0.482 1.500e+09 Logistics & Transportation 10800 4050 Advanced Disposal 4.51 1.400e+09 Logistics & Transportation 2116 406 Carabsoft Technology 1.42 1.400e+09 Government Services 3365 4141 Restoration Hardware 0.91 1.200e+09 Retail 2900 4050 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 4373 KSS 0.78 1.000e+09 Health 761 4373 KSS 0.78 1.000e+09 Manufacturing 8800 4532 Gobaddy.com 0.49 9.109e+08 Energy 492 4532 Gobaddy.com 0.49 9.109e+08 Energy 492 4533 Cobaddy.com 0.49 1.50e+08 Energy 492 4532 Gobaddy.com 0.41 8.615e+08 Business Products & Services 1257 4793 LORD Corporation 0.41 8.615e+08 Business Products & Services 1257 4794 Hearthside Food Solutions 3.17 8.396e+08 Energy 270 4795 Genesis-ATC 1.15 8.40e+08 Energy 270 4795 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 4896 Universal Services of America 0.40	4052	Kum & Go	0.65	2.800e+09	Retail	4589
2522 DLA Piper	4802	Boise Cascade	0.41	2.800e+09	Construction	4470
4629 Prime Therapeutics	1397	EnvisionRxOptions	2.88	2.700e+09	Health	625
A Bridger	2522	DLA Piper	1.41	2.400e+09	Business Products & Services	4036
1843 Sun Coast Resources 2.08 1.990e+09 Energy 1640 3844 Atlas Oil Company 0.74 1.990e+09 Logistics & Transportation 374 14961 Kirkland & Ellis 0.36 1.900e+09 Business Products & Services 1517 1488 Sprouts Farmers Market 2.68 1.800e+09 Consumer Products & Services 13200 4689 Global Brass and Copper Holdings 0.45 1.700e+09 Manufacturing 1986 198	4629	Prime Therapeutics	0.47	2.000e+09	Health	2549
3844 Atlas Oil Company	4		233.08	1.900e+09	Energy	50
4961 Kirkland & Ellis 0.36 1.900e+09 Business Products & Services 1517 1488 Sprouts Farmers Market 2.68 1.800e+09 Consumer Products & Services 13200 4689 Global Brass and Copper Holdings 0.45 1.700e+09 Manufacturing 1986 3463 Hogan Lovells 0.89 1.600e+09 Business Products & Services 2280 2145 AdvancePierre Foods 1.73 1.500e+09 Food & Beverage 4000 960 Genco 0.82 1.500e+09 Logistics & Transportation 10800 960 Advanced Disposal 4.51 1.400e+09 Environmental Services 5347 2236 Total Quality Logistics 1.65 1.400e+09 Logistics & Transportation 2116 2496 Carabsoft Technology 1.42 1.400e+09 Retail 2900 1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Retail 2900 1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Manufacturing 8500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 115 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quim Emanuel 1.03 8.525e+08 Business Products & Services 697 1284 Hearthside Food Solutions 3.17 8.36e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Business Products & Services 1257 427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4880 Goodman Networks 1.23 6.509e+08 Telecommunications 1000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Business Products & Services 1257 427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4009	1843	Sun Coast Resources	2.08	1.900e+09	Energy	1640
1488 Sprouts Farmers Market 2.68 1.800e+09 Consumer Products & Services 13200 4689 Global Brass and Copper Holdings 0.45 1.700e+09 Business Products & Services 2280 2145 AdvancePierre Foods 1.73 1.500e+09 Business Products & Services 2280 2145 AdvancePierre Foods 1.73 1.500e+09 Food & Beverage 4000 3650 Genco 0.82 1.500e+09 Eogistics & Transportation 10800 4000 2000 200000 20000 20000 20000 20000 20000 20000 20000 20	3844	Atlas Oil Company	0.74	1.900e+09	Logistics & Transportation	374
1986 Global Brass and Copper Holdings 0.45 1.700e+09 Manufacturing 1986 3463 Hogan Lovells 0.89 1.600e+09 Business Products & Services 2280 1.600e+09 Gence 1.73 1.500e+09 Food & Beverage 4000 3650 Genco 0.82 1.500e+09 Environmental Services 5347 1.400e+09 Environmental Services 5347 1.400e+09 Environmental Services 5347 1.400e+09 Environmental Services 5347 1.400e+09 Environmental Services 3.65 3414 Restoration Hardware 0.91 1.200e+09 Retail 2.900 1.908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Manufacturing 8.500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 492 4932 Golbaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 2.52 4793 LORD Corporation 0.41 8.615e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Energy 5000 2.47 4860 Energy 5000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Business Products & Services 748 4860 4870 48	4961	Kirkland & Ellis	0.36	1.900e+09	Business Products & Services	1517
3463 Hogan Lovells	1488	Sprouts Farmers Market	2.68	1.800e + 09	Consumer Products & Services	13200
2145	4689	Global Brass and Copper Holdings	0.45	1.700e + 09	Manufacturing	1986
3650 Genco 0.82 1.500e+09 Logistics & Transportation 10800 960 Advanced Disposal 4.51 1.400e+09 Environmental Services 5347 2236 Total Quality Logistics 1.65 1.400e+09 Logistics & Transportation 2116 2496 Carahsoft Technology 1.42 1.400e+09 Logistics & Transportation 2116 3414 Restoration Hardware 0.91 1.200e+09 Retail 2900 1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Heanth 761 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 </td <td>3463</td> <td>Hogan Lovells</td> <td>0.89</td> <td>1.600e + 09</td> <td>Business Products & Services</td> <td>2280</td>	3463	Hogan Lovells	0.89	1.600e + 09	Business Products & Services	2280
960 Advanced Disposal 4.51 1.400e+09 Environmental Services 5347 2236 Total Quality Logistics 1.65 1.400e+09 Logistics & Transportation 2116 2496 Carahsoft Technology 1.42 1.400e+09 Government Services 365 3414 Restoration Hardware 0.91 1.200e+09 Retail 2900 1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Manufacturing 8500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 2961 Genesis-ATC 1.15	2145	AdvancePierre Foods	1.73	1.500e + 09	Food & Beverage	4000
2236 Total Quality Logistics 1.65 1.400e+09 Logistics & Transportation 2116 2496 Carahsoft Technology 1.42 1.400e+09 Government Services 365 3414 Restoration Hardware 0.91 1.200e+09 Retail 2900 1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Manufacturing 8500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.52e+08 Buiness Products & Services 697 2961 Genesis-ATC 1.15	3650	Genco	0.82	1.500e + 09	Logistics & Transportation	10800
2496 Carahsoft Technology 1.42 1.400e+09 Government Services 365 3414 Restoration Hardware 0.91 1.200e+09 Retail 2900 1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Manufacturing 8500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 <	960	Advanced Disposal	4.51	1.400e + 09	Environmental Services	5347
2496 Carahsoft Technology 1.42 1.400e+09 Government Services 365 3414 Restoration Hardware 0.91 1.200e+09 Retail 2900 1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Manufacturing 8500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 <	2236	Total Quality Logistics	1.65	1.400e + 09	Logistics & Transportation	2116
1908 Diplomat Specialty Pharmacy 1.99 1.100e+09 Health 761 3734 KSS 0.78 1.000e+09 Manufacturing 8500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 409 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	2496		1.42	1.400e+09		365
3734 KSS 0.78 1.000e+09 Manufacturing 8500 3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications	3414	Restoration Hardware	0.91	1.200e+09	Retail	2900
3425 Blackhawk Network Holdings 0.90 9.591e+08 Financial Services 725 1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold	1908	Diplomat Specialty Pharmacy	1.99	1.100e+09	Health	761
1996 Ambit Energy 1.88 9.303e+08 Energy 492 4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Univers	3734	KSS	0.78	1.000e+09	Manufacturing	8500
4532 GoDaddy.com 0.49 9.109e+08 IT Services 3369 2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 <td>3425</td> <td>Blackhawk Network Holdings</td> <td>0.90</td> <td>9.591e + 08</td> <td>Financial Services</td> <td>725</td>	3425	Blackhawk Network Holdings	0.90	9.591e + 08	Financial Services	725
2958 ImmixGroup 1.15 8.836e+08 IT Services 252 4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427	1996	Ambit Energy	1.88	9.303e + 08	Energy	492
4793 LORD Corporation 0.41 8.615e+08 Manufacturing 2959 3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 459<	4532	GoDaddy.com	0.49	9.109e + 08	IT Services	3369
3163 Quinn Emanuel 1.03 8.525e+08 Business Products & Services 697 2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4559 Belcan 0.52 6.887e+08 Engineering 10000 4806	2958	ImmixGroup	1.15	8.836e + 08	IT Services	252
2961 Genesis-ATC 1.15 8.462e+08 Telecommunications 347 1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Health 1945 4815 Perkins Coi	4793	LORD Corporation	0.41	8.615e + 08	Manufacturing	2959
1284 Hearthside Food Solutions 3.17 8.396e+08 Food & Beverage 5000 1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perki	3163	Quinn Emanuel	1.03	8.525e + 08	Business Products & Services	697
1480 Coyote Logistics 2.70 7.864e+08 Logistics & Transportation 1219 4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Ce	2961	Genesis-ATC	1.15	8.462e+08	Telecommunications	347
4765 Squire Sanders 0.42 7.745e+08 Business Products & Services 1257 3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair	1284	Hearthside Food Solutions	3.17	8.396e + 08	Food & Beverage	5000
3921 Granite Telecommunications 0.70 7.362e+08 Telecommunications 1000 4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services	1480	Coyote Logistics	2.70	7.864e + 08	Logistics & Transportation	1219
4854 Arnold & Porter 0.40 7.310e+08 Business Products & Services 748 1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy <td>4765</td> <td>Squire Sanders</td> <td>0.42</td> <td>7.745e + 08</td> <td>Business Products & Services</td> <td>1257</td>	4765	Squire Sanders	0.42	7.745e + 08	Business Products & Services	1257
1869 Universal Services of America 2.04 7.181e+08 Security 20000 2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	3921	Granite Telecommunications	0.70	7.362e + 08	Telecommunications	1000
2274 Liberty Power 1.61 7.108e+08 Energy 277 4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	4854	Arnold & Porter	0.40	7.310e + 08	Business Products & Services	748
4427 Sunshine Minting 0.53 7.061e+08 Manufacturing 280 4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	1869	Universal Services of America	2.04	7.181e + 08	Security	20000
4459 Belcan 0.52 6.887e+08 Engineering 10000 2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	2274	Liberty Power	1.61	7.108e + 08	Energy	277
2806 Goodman Networks 1.23 6.509e+08 Telecommunications 1693 4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	4427	Sunshine Minting	0.53	7.061e + 08	Manufacturing	280
4009 Schumacher Group 0.67 6.130e+08 Health 1945 4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	4459	Belcan	0.52	6.887e + 08	Engineering	10000
4815 Perkins Coie 0.40 6.080e+08 Business Products & Services 823 1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	2806	Goodman Networks	1.23	6.509e + 08	Telecommunications	
1768 The Cellular Connection 2.17 6.065e+08 Telecommunications 1428 2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	4009	Schumacher Group	0.67	6.130e + 08	Health	1945
2537 Wayfair.com 1.40 6.020e+08 Retail 1300 4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	4815	Perkins Coie	0.40	6.080e + 08	Business Products & Services	823
4577 Sutherland Global Services 0.48 5.976e+08 Business Products & Services 32000 4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75	1768	The Cellular Connection	2.17	6.065e + 08	Telecommunications	1428
4093 Advanced BioEnergy 0.64 5.848e+08 Energy 75			1.40	6.020e + 08		1300
	4577	Sutherland Global Services	0.48	5.976e + 08	Business Products & Services	32000
4857 AVI-SPL 0.39 5.807e+08 Business Products & Services 1800		Advanced BioEnergy	0.64	5.848e + 08		75
	4857	AVI-SPL	0.39	5.807e + 08	Business Products & Services	1800

```
## 2
                                       6549
                                                               WI 4.7 Billion
                     Construction
                                                     Beloit
## 3 Consumer Products & Services
                                      10000
                                                   New York
                                                               NY 4.6 Billion
## 4
                                       3919
                                               Mt. Sterling
                                                               IL 4.5 Billion
                  Food & Beverage
## 5
                      IT Services
                                       3000
                                                  Tarrytown
                                                               NY 3.8 Billion
## 6 Consumer Products & Services
                                       3341
                                               Huntersville
                                                               NC 3.5 Billion
                                                               IA 2.8 Billion
## 7
                           Retail
                                       4589 West Des Moines
## 8
                     Construction
                                       4470
                                                      Boise
                                                               ID 2.8 Billion
                                                               OH 2.7 Billion
## 9
                                        625
                           Health
                                                  Twinsburg
## 10 Business Products & Services
                                       4036
                                                    Chicago
                                                               IL 2.4 Billion
unique(inc$Industry)
## [1] "Consumer Products & Services" "Government Services"
## [3] "Health"
                                      "Energy"
## [5] "Advertising & Marketing"
                                      "Real Estate"
   [7] "Financial Services"
                                      "Retail"
## [9] "Software"
                                      "Computer Hardware"
## [11] "Logistics & Transportation"
                                      "Food & Beverage"
## [13] "IT Services"
                                      "Business Products & Services"
## [15] "Education"
                                      "Construction"
## [17] "Manufacturing"
                                      "Telecommunications"
## [19] "Security"
                                      "Human Resources"
## [21] "Travel & Hospitality"
                                      "Media"
## [23] "Environmental Services"
                                      "Engineering"
## [25] "Insurance"
unique(sample$Industry)
## [1] "Computer Hardware"
                                      "Construction"
## [3] "Consumer Products & Services" "Food & Beverage"
## [5] "IT Services"
## [7] "Health"
                                     "Business Products & Services"
#Plot to show by State, industries with the highest revenue where employees > 1000 from Sample of 50
ggplot(filter(sample, Employees > 1000),aes(x = State, y = Revenue, color = Industry)) +
 geom point() +
 geom_label(aes(label = Industry)) +
 labs(x = "State", y = "Revenue") +
  theme(legend.position = "none")
```



Question 1

Create a graph that shows the distribution of companies in the dataset by State (ie how many are in each state). There are a lot of States, so consider which axis you should use. This visualization is ultimately going to be consumed on a 'portrait' oriented screen (ie taller than wide), which should further guide your layout choices.

Answer 1

I used ggplot below to show a portrait graph by using coord_flip(). Then using Theme_Tufte. Additionally then I tried to plot the numbers on the bar graph and adjusted to size 3 so I can show the numbers on the graph.

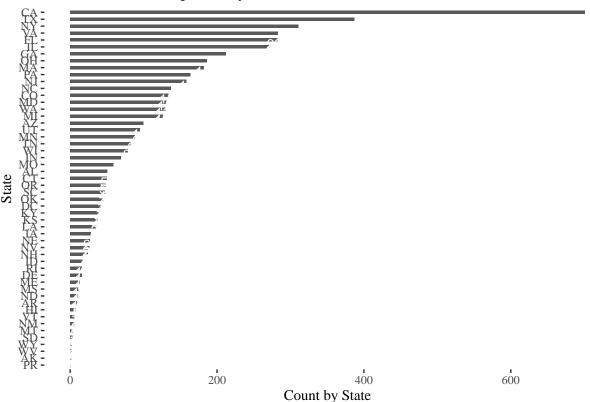
```
# Answer Question 1 here

library(ggplot2)
library(tidyverse)
library(dplyr)
library(scales)
```

```
##
## Attaching package: 'scales'
```

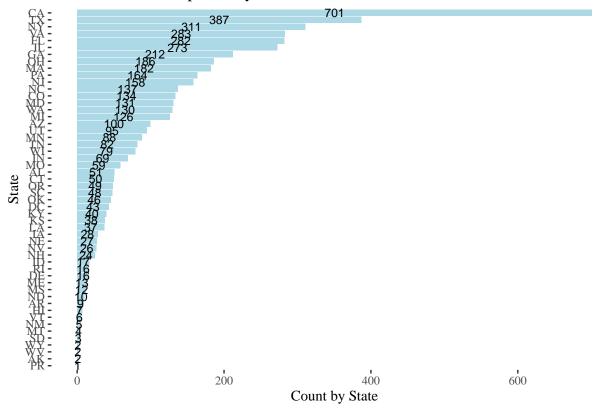
```
## The following object is masked from 'package:purrr':
##
##
       discard
## The following object is masked from 'package:readr':
##
##
       col_factor
#qraph1
inc %>% count(State) %>%
  ggplot(aes(x=reorder(State, n), y=n)) +
 geom_bar(stat = 'identity', width = 0.5, position = position_dodge(0.8)) +
  geom_text(aes(label=n), vjust=1.6, color="white", size=3.5) +
 coord_flip() +
 theme_tufte() +
 xlab("State") +
 ylab("Count by State") +
  ggtitle("Distribution of companies by state")
```

Distribution of companies by state



```
coord_flip() +
theme_tufte() +
xlab("State") +
ylab("Count by State") +
ggtitle("Distribution of companies by state") + scale_color_brewer(palette = "Pastel1")
```

Distribution of companies by state



Question 2

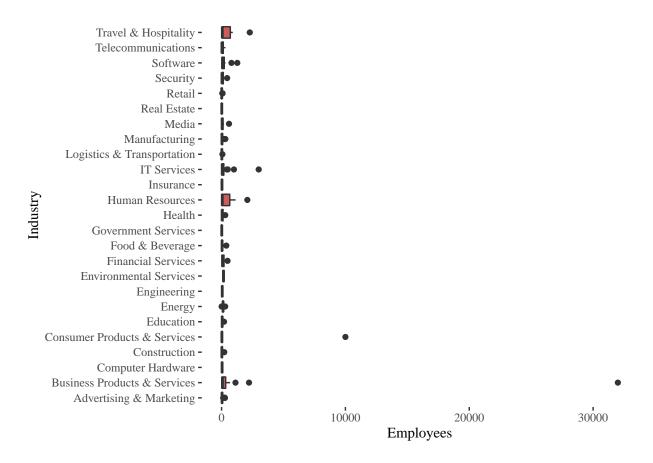
Lets dig in on the state with the 3rd most companies in the data set. Imagine you work for the state and are interested in how many people are employed by companies in different industries. Create a plot that shows the average and/or median employment by industry for companies in this state (only use cases with full data, use R's complete.cases() function.) In addition to this, your graph should show how variable the ranges are, and you should deal with outliers.

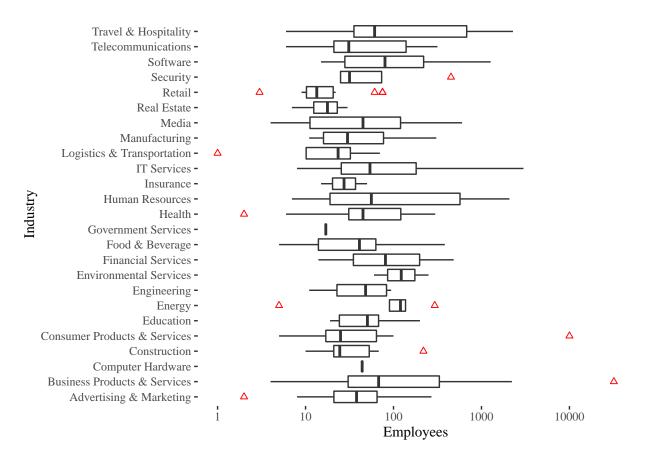
Answer 2:

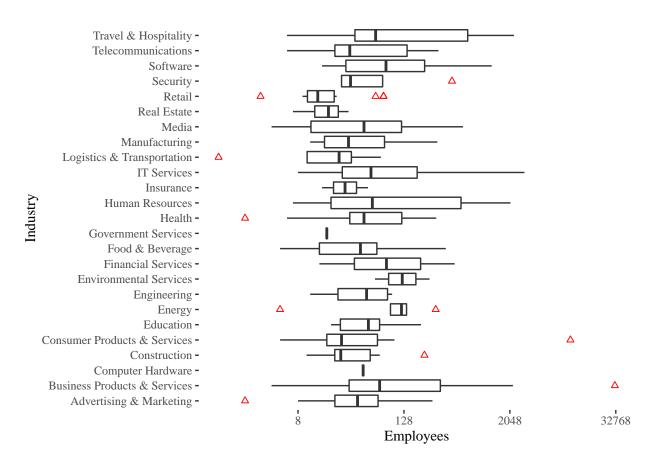
I tried to create a few plots to understand how to show the outlier data using axis transformation for log2 and log10. Also showed triangular on outlier data. User reorder method The "default" method treats its first argument as a categorical variable (Industry), and reorders its levels based on the values of a second variable (Employees), usually numeric

```
# Answer Question 2 here
#1 Create a plot that shows the average and/or median employment by industry for companies in this stat
# Subset only NY records
NyState <- inc[ which(inc$State == 'NY'), ]</pre>
# Remove incomplete records
NyState <- NyState[complete.cases(NyState), ]</pre>
# show data
head(NyState )
##
      Rank
                                Name Growth_Rate Revenue
## 26
                        BeenVerified
                                           84.43 13700000
       26
## 30
       30
                            Sailthru
                                           73.22 8100000
## 37
       37
                        YellowHammer
                                           67.40 18000000
## 38
       38
                           Conductor
                                           67.02 7100000
       48 Cinium Financial Services
                                           53.65 5900000
## 48
## 70
       70
                                           44.99 27900000
                            33Across
##
                          Industry Employees
                                                  City State
## 26 Consumer Products & Services
                                          17 New York
## 30
          Advertising & Marketing
                                         79 New York
## 37
          Advertising & Marketing
                                         27 New York
                                                         NY
## 38
          Advertising & Marketing
                                         89 New York
                                                         NY
## 48
               Financial Services
                                          32 Rock Hill
                                                         NY
## 70
          Advertising & Marketing
                                         75 New York
                                                         NY
#Find statistics
summary(NyState)
##
        Rank
                       Name
                                      Growth_Rate
                                                         Revenue
## Min. : 26
                  Length:311
                                     Min. : 0.350
                                                              :2.000e+06
                                                       Min.
                                                       1st Qu.:4.300e+06
  1st Qu.:1186
                  Class :character
                                     1st Qu.: 0.670
## Median :2702
                  Mode :character
                                     Median : 1.310
                                                      Median :8.800e+06
## Mean
         :2612
                                     Mean
                                            : 4.371
                                                       Mean
                                                             :5.872e+07
   3rd Qu.:4005
                                      3rd Qu.: 3.580
                                                       3rd Qu.:2.570e+07
##
          :4981
## Max.
                                     Max.
                                             :84.430
                                                       Max. :4.600e+09
##
      Industry
                         Employees
                                             City
                                                               State
## Length:311
                       Min.
                             :
                                  1.0
                                        Length:311
                                                           Length:311
## Class :character
                       1st Qu.:
                                  21.0
                                        Class : character Class : character
                                        Mode :character Mode :character
## Mode :character
                       Median :
                                 45.0
##
                       Mean
                            : 271.3
##
                       3rd Qu.: 105.5
##
                       Max.
                              :32000.0
NyState %% arrange(desc(Employees)) %% head(50) %% kable() %% kable_styling()
#quick graph of NY state
ggplot(NyState, aes(x = Industry, y = Employees)) +
  geom_boxplot(fill = "indianred") +
  labs(x = "Industry", y = "Employees") +
  theme_tufte() + coord_flip()
```

Rank	Name	Growth Rate	Revenue	Industry
$\frac{4577}{}$	Sutherland Global Services	0.48	5.976e + 08	Business Products & Services
4936	Coty	0.36	4.600e + 09	Consumer Products & Services
4716	Westcon Group	0.44	3.800e + 09	IT Services
3899	Denihan Hospitality Group	0.71	2.808e + 08	Travel & Hospitality
4363	TransPerfect	0.55	3.413e + 08	Business Products & Services
1499	Sterling Infosystems	2.66	2.149e + 08	Human Resources
4465	OpenLink	0.52	3.043e + 08	Software
3136	FSO Onsite Outsourcing	1.05	5.530e + 07	Human Resources
2830	ReSource Pro	1.22	2.430e + 07	Business Products & Services
2995	Pride Technologies	1.13	2.310e + 08	Human Resources
$\frac{2385}{3387}$	Mitchell/Martin	0.92	1.453e + 08	IT Services
4747	TravelClick	0.43	2.571e + 08	Travel & Hospitality
2964	DataArt	1.14	3.070e+07	Software
4913	Jackson Lewis	0.37	3.520e + 08	Business Products & Services
$\frac{4646}{4646}$	Mimeo.com	0.46	8.770e + 07	Business Products & Services
$\frac{4040}{4153}$	Everyday Health	0.62	1.464e + 08	Media
$\frac{4103}{4003}$	Ovation Travel Group	0.67	5.860e + 07	Travel & Hospitality
$\frac{4003}{1640}$	BlueWolf	2.38	9.040e+07	IT Services
$\frac{1040}{2971}$	ConServe	1.14	5.270e+07	Financial Services
$\frac{2371}{4727}$	Arrow Security	0.44	1.400e+07	Security
$\frac{4727}{3584}$	Infusion	0.44	6.790e+07	IT Services
$\frac{-3534}{4535}$	Capital Access Network	0.49	1.516e + 08	Financial Services
$\frac{4535}{4224}$	Fragomen	0.49	3.465e+08	Business Products & Services
$\frac{4224}{3500}$	Magnolia Bakery	0.88	2.360e+08	Food & Beverage
$\frac{3500}{1069}$	Systems Made Simple	3.94	1.671e + 08	IT Services
$\frac{1009}{4039}$	Empire Office	0.66	3.496e + 08	Business Products & Services
$\frac{4039}{3704}$	Infinity Consulting Solutions	0.79	3.490e+08 3.570e+07	Human Resources
$\frac{3704}{3021}$	Linium	1.11	3.570e + 07 4.590e + 07	IT Services
$\frac{3021}{2218}$	Globo Mobile	1.11	4.590e+07 4.500e+06	Software
$\frac{2218}{2896}$	Telx	1.18	4.500e+00 2.143e+08	Telecommunications
$\frac{2890}{4820}$	Aluf Plastics	0.40		Manufacturing Manufacturing
		1.48	1.029e + 08	IT Services
2436	Shinetech Software		1.200e+07	
3871	eTransMedia Technology	0.72	2.590e+07	Health
4981	SmartSource Computer & Audio Visual Rentals	0.35	5.730e + 07	Business Products & Services
1190	Usablenet	3.52	5.900e+07	IT Services
2556	Precision Pipeline Solutions	1.39	3.120e+07	Energy
4944	McElroy Deutsch	0.36	1.165e + 08	Business Products & Services
4732	Mycroft	0.43	2.640e+07	IT Services
4552	Paradysz /TMC	0.49	4.100e+07	Advertising & Marketing
4738	McMurry/TMG	0.43	9.140e+07	Advertising & Marketing
3009	5Linx Enterprises	1.12	1.036e + 08	Telecommunications
1565	Droga5	2.54	6.730e + 07	Advertising & Marketing
2997	Fibertech Networks	1.13	1.423e + 08	Telecommunications
3661	Environmental Products & Services of Vermont	0.81	4.510e+07	Environmental Services
1895	nfrastructure	2.00	5.320e+07	IT Services
1371	Impelsys	2.94	1.290e+07	Software
3054	DDS Companies	1.09	5.250e + 07	Construction
275	Technical Solutions	15.85	1.630e + 07	Telecommunications
3262	Sriven Systems	0.98	2.530e+07	IT Services
890	Payoneer	5.00	4.040e + 07	Financial Services







```
#FINAL GRAPH FOR QUESTION2: Graph to show boxplot with log2 transformation. Use re-order. Shows Busin
ggplot(NyState, aes(reorder(Industry, Employees, FUN = median), Employees)) +
  geom_boxplot(outlier.colour = "Red",
              outlier.shape = 2) +
 scale_y_continuous(trans = "log2", limits = c(min(NyState$Employees), max(NyState$Employees))) +
  coord flip() +
  geom_hline(yintercept = median(NyState$Employees),
             color="Red",
             linetype="dashed") +
  geom_text(aes(x=2,
            label="Median for Employees", y = median(NyState$Employees)+100 ),
            size = 3,
            colour="blue") +
  xlab("Industry") +
  ylab("Number of Employees") +
  theme_tufte() +
  ggtitle("Number of Employees by Industry in New York")
```

Warning: Use of `NyState\$Employees` is discouraged. Use `Employees` instead.

Number of Employees by Industry in New York



Question 3

Now imagine you work for an investor and want to see which industries generate the most revenue per employee. Create a chart that makes this information clear. Once again, the distribution per industry should be shown.

Answer 3:

xlab("Industry") +

ylab("Revenue per employee (\$)") +

I checked the summary by using table and then colored the graph light blue to keep it more easy to view. Highest revenue is Computer Hardware, which generates revenue and employeers fewer total employees within this industry.

```
inc_nonmissing <- inc[complete.cases(inc), ]
revemployee <- group_by(inc_nonmissing, Industry) %>% summarise(sum(Revenue),sum(Employees), revperemp

revemployee %>% arrange(desc(revperemp)) %>% head(50) %>% kable() %>% kable_styling()

ggplot(revemployee, aes(x=reorder(Industry, revperemp), y=revperemp)) +
    geom_bar(stat = 'Identity',fill = "light blue", position = "dodge") +
    coord_flip() +
```

Industry	sum(Revenue)	sum(Employees)	revperemp
Computer Hardware	11885700000	9714	1223563.93
Energy	13771600000	26437	520921.44
Construction	13174300000	29099	452740.64
Logistics & Transportation	14837800000	39994	371000.65
Consumer Products & Services	14956400000	45464	328972.37
Insurance	2337900000	7339	318558.39
Manufacturing	12603600000	43942	286823.54
Retail	10257400000	37068	276718.46
Financial Services	13150900000	47693	275740.67
Environmental Services	2638800000	10155	259852.29
Telecommunications	7287900000	30842	236297.91
Government Services	6009100000	26185	229486.35
Business Products & Services	26345900000	117357	224493.64
Health	17860100000	82430	216669.90
IT Services	20525000000	102788	199682.84
Advertising & Marketing	7785000000	39731	195942.71
Food & Beverage	12812500000	65911	194390.92
Media	1742400000	9532	182794.80
Software	8134600000	51262	158686.75
Real Estate	2956800000	18893	156502.41
Education	1139300000	7685	148249.84
Travel & Hospitality	2931600000	23035	127267.20
Engineering	2532500000	20435	123929.53
Security	3812800000	41059	92861.49
Human Resources	9246100000	226980	40735.31

```
ggtitle("Revenue per employee for industry") +
scale_y_continuous(labels = scales::comma)+
theme_tufte()
```

Revenue per employee for industry

