R Notebook

```r
library(tidyverse)
library(plotrix)
library(ggstatsplot)
library(ggridges)
```

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:

```r
# import data from github
inc <- read.csv("https://raw.githubusercontent.com/charleyferrari/CUNY_DATA_608/master/module1/Data/inc5
```

And lets preview this data:

```r
# preview 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
```

```r
# view data summary
summary(inc)
```

```
##      Rank          Name            Growth_Rate         Revenue
##  Min.   :   1   Length:5001        Min.   :  0.340   Min.   :2.000e+06
##  1st Qu.:1252   Class :character   1st Qu.:  0.770   1st Qu.:5.100e+06
##  Median :2502   Mode  :character   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
##  Max.   :5000                      Max.   :421.480   Max.   :1.010e+10
##
```

```
##    Industry          Employees          City            State
##  Length:5001      Min.   :   1.0   Length:5001      Length:5001
##  Class :character  1st Qu.:   25.0  Class :character  Class :character
##  Mode  :character  Median :   53.0  Mode  :character  Mode  :character
##                    Mean   :  232.7
##                    3rd Qu.:  132.0
##                    Max.   :66803.0
##                    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:

Ok, for me its easier to write out and define what they mean:

- `Rank` is the rank of the company in regard to `Growth Rate` (ordinal)
- `Name` is the name of the company
- `Growth Rate` is the rate at which the company has grown over a period of time (continuous).
- `Revenue` is the amount of money each company has made (continuous)
- `Industry` is a categorical variable stating the industry type of the company
- `Employees` are the number of employees (continuous)
- `City` is the city where the company is located
- `State` is the state where the company is located

So for any object data types the `summary()` function is simply counting the records and defining the class type as `character`. For numeric data types we are getting some summary statistics including: - Min - 1st Quartile - Median - Mean - 3rd Quartile - Max - How many NA's

Some grouping by categorical data would help me better understand where companies that have high `Growth Rate` tend to be located.

```
# Insert your code here, create more chunks as necessary
(top_5_states <- inc %>%
  group_by(State) %>%
  tally(sort = T) %>%
  head())
```

```
## # A tibble: 6 x 2
##   State     n
##   <chr> <int>
## 1 CA      701
## 2 TX      387
## 3 NY      311
## 4 VA      283
## 5 FL      282
## 6 IL      273
```

```
# top 5 states with most companies and average growth rate
inc %>%
  group_by(State) %>%
  summarise(n = n(), mean_growth_rate = round(mean(Growth_Rate),2)) %>%
  arrange(desc(n)) %>%
  head() %>%
  arrange(desc(mean_growth_rate))
```

```
## # A tibble: 6 x 3
##    State     n mean_growth_rate
##    <chr> <int>            <dbl>
## 1 TX      387             6.02
## 2 CA      701             5.9
## 3 FL      282             5.85
## 4 VA      283             4.88
## 5 NY      311             4.37
## 6 IL      273             3.74
```

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.

```r
# Answer Question 1 here
inc %>%
  group_by(State) %>%
  summarise(count = n()) %>%
  ggplot() +
    geom_bar(aes(count, reorder(State, count), fill = count),
             stat = "identity") +
    scale_fill_gradient(low = "white", high = "darkblue") +
    labs(x="Count", y="State",
         title = "Distribution of Companies by State") +
    theme(plot.title.position = "plot",
          plot.title = element_text(hjust = 0.5,
                                    size = 18),
          axis.text.y = element_text(size = 6),
          axis.title.y = element_text(size = 12),
          legend.position = "none")
```

Distribution of Companies by State
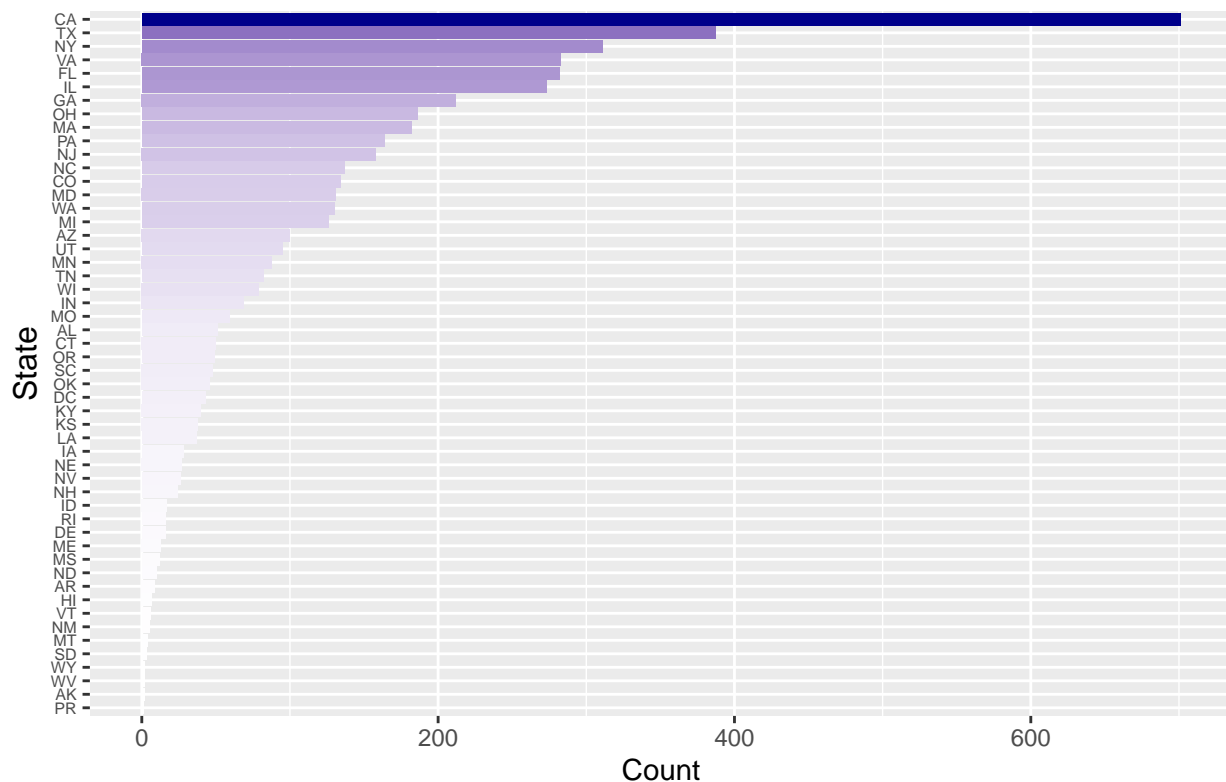
Question 2

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 Question 2 here
# isloate state with 3rd most highest companies (NY)
inc_ny <- inc[inc$State == "NY",]

# look at summary for NY employees column
summary(inc_ny$Employees)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0    21.0    45.0   271.3   105.5 32000.0
```

Let's visualize `Employees` in a boxplot

```
inc_ny %>%
  ggplot() +
    geom_boxplot(aes(Employees)) +
    coord_flip()
```
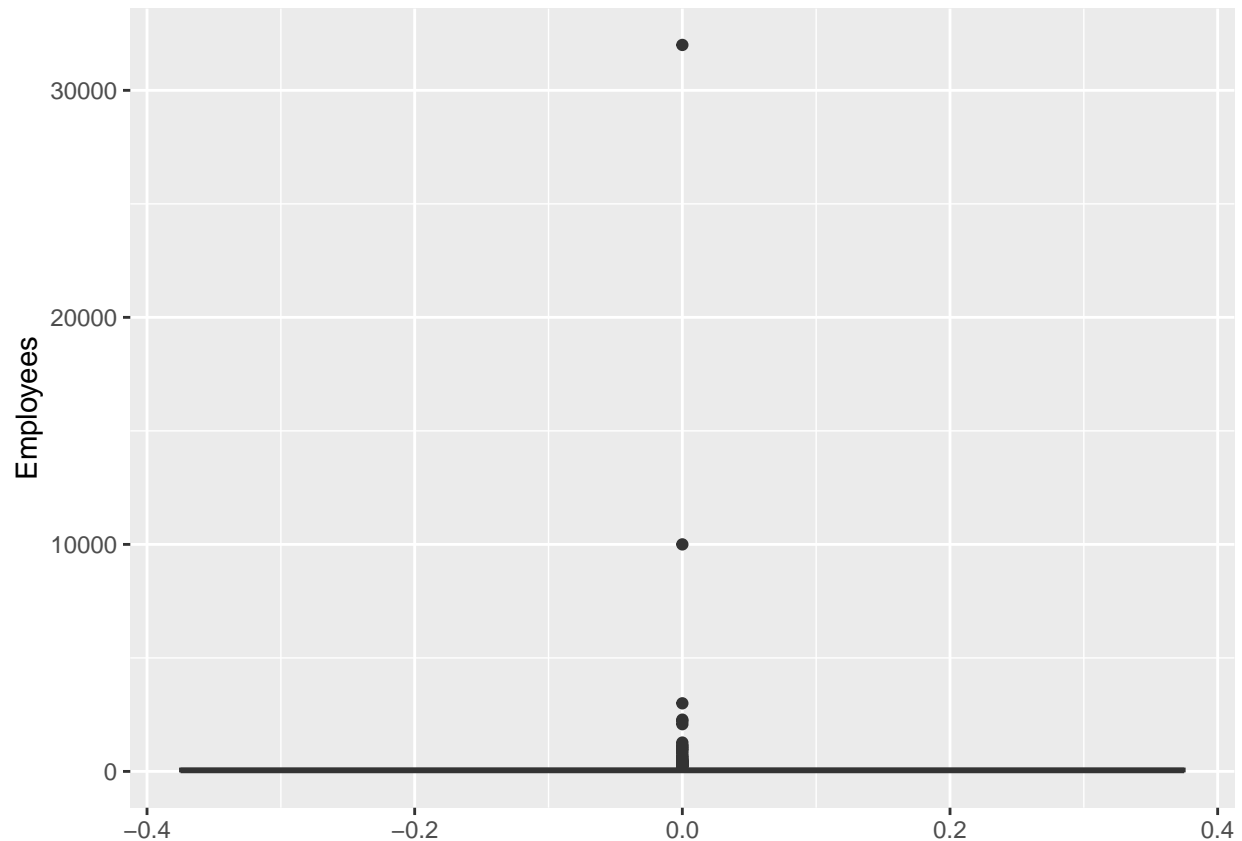
Looks like we have some extreme variation, wonder what industries this is in.

Let's group by industry and create min, max, mean and median columns sorted decreasing by max.

First let's create custom function.

```r
# create group by function for particular column
# aggregates by another and sorts descending by another (default max)
groupby_column <- function(df, col_grp, agg_col, sort_col=max_grp) {

  # quote columns
  col_grp <- dplyr::enquo(col_grp)
  agg_col <- dplyr::enquo(agg_col)
  sort_col <- dplyr::enquo(sort_col)

  # groupby and add aggreations
  result <- df %>%
    dplyr::group_by(!!col_grp) %>%
    dplyr::summarise(n_grp = n(),
              min_grp = min(!!agg_col),
              max_grp = max(!!agg_col),
              mean_grp = round(mean(!!agg_col),2),
              med_grp = round(median(!!agg_col),2)) %>%
    arrange(desc(!!sort_col))

  return(result)
}
```

Let's apply the custom function.

```r
# apply function to dataset groupby Industry, aggregated by Employees
(inc_ny_grouped <- groupby_column(df = inc_ny,
                                  col_grp = Industry,
                                  agg_col = Employees))
```

```
## # A tibble: 25 x 6
##    Industry                    n_grp min_grp max_grp mean_grp med_grp
##    <chr>                       <int>   <int>   <int>    <dbl>   <dbl>
##  1 Business Products & Services   26       4   32000    1492.    70.5
##  2 Consumer Products & Services   17       5   10000     626.    25
##  3 IT Services                    43       8    3000     204.    54
##  4 Travel & Hospitality            7       6    2280     548.    61
##  5 Human Resources                11       7    2081     438.    56
##  6 Software                       13      15    1271     246.    80
##  7 Media                          11       4     602     108     45
##  8 Financial Services             13      14     483     144.    81
##  9 Security                        4      25     450     135     32.5
## 10 Food & Beverage                 9       5     383      76.4   41
## # ... with 15 more rows
```

Extreme outliers in `Business Products & Services` and `Consumer Products & Services` will skew any visualizations. Try as is first.

Try with `geom_crossbar()`

First create custom function.

```r
# create custom function to plot with geom_crossbar
# enter df, x, y, sort order for x, ymin, ymax and fill (default grey)
crossbar_plot <- function(df, x, y, sort_x, ymin, ymax,
                          fill ="grey", xlab="", ylab="Range",
                          title = "", width = 0.5) {
  # quote columns
  x <- dplyr::enquo(x)
  y <- dplyr::enquo(y)
  sort_x <- dplyr::enquo(sort_x)
  ymin <- dplyr::enquo(ymin)
  ymax <- dplyr::enquo(ymax)

  # plot
  result <- df %>%
    ggplot(aes(x = reorder(as.factor(!!x), !!sort_x), y = !!y)) +
      geom_crossbar(aes(ymin = !!ymin, ymax = !!ymax), width = width,
                fill = fill) + coord_flip() +
      labs(x = xlab, y=ylab, title = title)

  return(result)
}
```

Let's try out custom function.

```
# plot by Median Employee by Industry
crossbar_plot(inc_ny_grouped, Industry, med_grp, med_grp, min_grp, max_grp,
              title = "Median Employee by Industry")
```


Median Employee by Industry
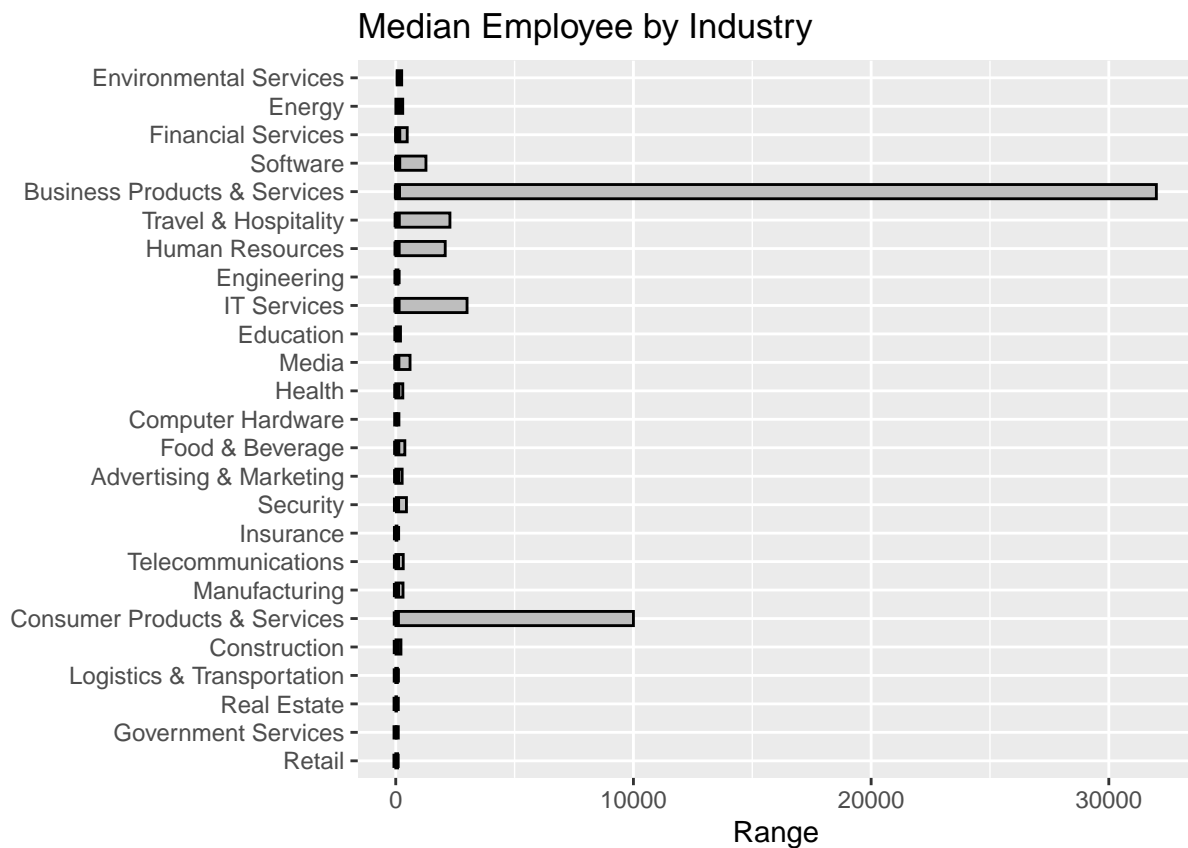
Doesn't deal with outliers well, impossible to tell what actual median is. This is NOT a good visualation.

Try Log transform and re-plot.

Transformation

```
# log10 transform Employee column
(inc_ny_emp_log <- inc_ny %>%
  mutate(emp_log = round(log10(Employees)),2) %>%
  select(Rank, Name, Employees, emp_log, everything()) %>%
  arrange(desc(emp_log)))
```

```
##      Rank                          Name Employees emp_log
## 1    4577    Sutherland Global Services     32000       5
## 2    4936                          Coty     10000       4
## 3    1069           Systems Made Simple       382       3
## 4    1499          Sterling Infosystems      2081       3
## 5    1640                      BlueWolf       500       3
## 6    2218                  Globo Mobile       320       3
## 7    2830                   ReSource Pro      1124       3
```

7

```
## 8   2964                        DataArt    800   3
## 9   2971                        ConServe    483   3
## 10  2995               Pride Technologies   1000   3
## 11  3021                          Linium    322   3
## 12  3136           FSO Onsite Outsourcing   1134   3
## 13  3387                  Mitchell/Martin   1000   3
## 14  3500                  Magnolia Bakery    383   3
## 15  3584                         Infusion    432   3
## 16  3704     Infinity Consulting Solutions    325   3
## 17  3899        Denihan Hospitality Group   2280   3
## 18  4003              Ovation Travel Group    509   3
## 19  4039                    Empire Office    345   3
## 20  4153                  Everyday Health    602   3
## 21  4224                         Fragomen    404   3
## 22  4363                     TransPerfect   2218   3
## 23  4465                         OpenLink   1271   3
## 24  4535            Capital Access Network    425   3
## 25  4646                       Mimeo.com    632   3
## 26  4716                    Westcon Group   3000   3
## 27  4727                  Arrow Security    450   3
## 28  4747                      TravelClick    906   3
## 29  4913                    Jackson Lewis    690   3
## 30    30                         Sailthru     79   2
## 31    38                        Conductor     89   2
## 32    48         Cinium Financial Services     32   2
## 33    70                         33Across     75   2
## 34    71                       LiveIntent     42   2
## 35   153                      Regal Wings     42   2
## 36   174                       Refinery29     99   2
## 37   219               Novisync Solutions    119   2
## 38   275              Technical Solutions    213   2
## 39   285                              SET     46   2
## 40   314                          Panjiva     48   2
## 41   450                  Carrot Creative     45   2
## 42   497                      RCS Capital    198   2
## 43   521              Happy Family Brands     51   2
## 44   522             Rockhedge Herb Farms     56   2
## 45   609                 SmartWatt Energy    137   2
## 46   656              ForwardThink Group     70   2
## 47   660                           OnDeck    153   2
## 48   661              Quality In Real Time    121   2
## 49   669                       Touchstorm     46   2
## 50   701            Thrillist Media Group    200   2
## 51   717                 Merchant Industry    102   2
## 52   750                       Livestream    146   2
## 53   752                      Private Prep    100   2
## 54   815                       Fingerpaint     51   2
## 55   830              Launchpad Advertising     40   2
## 56   883      Envisage Information Systems    142   2
## 57   890                         Payoneer    207   2
## 58   922                   Rebecca Minkoff     75   2
## 59   935                            Ioxus     90   2
## 60  1036                           Enstoa     40   2
## 61  1048                      Curbed Media     45   2
```

```
## 62  1054                             Tarte   100    2
## 63  1091               Blue Fountain Media   180    2
## 64  1119                          Adafruit    40    2
## 65  1129                     Likeable Media    65    2
## 66  1190                         Usablenet   295    2
## 67  1259                    Decorplanet.com    61    2
## 68  1293                      Madison Logic    51    2
## 69  1371                          Impelsys   220    2
## 70  1382                   City Sounds of NY    40    2
## 71  1423                       Media6Degrees    87    2
## 72  1442               Global Warranty Group    50    2
## 73  1451           Professional Data Systems    37    2
## 74  1503                  AdCorp Media Group    38    2
## 75  1565                            Droga5   250    2
## 76  1609                     Enter:Marketing    35    2
## 77  1706                                MKG    41    2
## 78  1755           Deep Blue Communications    47    2
## 79  1876                        Servomation    41    2
## 80  1895                      nfrastructure   237    2
## 81  1982                   Imagine Swimming   100    2
## 82  1985            U.S. Energy Development   120    2
## 83  2067             Signature Systems Group    77    2
## 84  2116 Charter School Business Management    33    2
## 85  2142             Liberty Jet Management    61    2
## 86  2173                     PM Pediatrics   181    2
## 87  2179                  Christmas Central    75    2
## 88  2198                          KJT Group    32    2
## 89  2209                               inVNT    44    2
## 90  2280                Celerant Technology    80    2
## 91  2336                         BountyJobs    56    2
## 92  2372           DiGennaro Communications    32    2
## 93  2399                               Rain    74    2
## 94  2405                           Sullivan    63    2
## 95  2436                 Shinetech Software   300    2
## 96  2475          Systems Management/Planning    52    2
## 97  2494                               RVM   106    2
## 98  2506                             QueBIT    46    2
## 99  2526              Soleo Communications    85    2
## 100 2542                   The Fortus Group    50    2
## 101 2556         Precision Pipeline Solutions   294    2
## 102 2678                    Treo Solutions   166    2
## 103 2718                        CreativeFeed    32    2
## 104 2751                            IOMedia   133    2
## 105 2795                            Onco360    67    2
## 106 2805             LeadDog Marketing Group   131    2
## 107 2828                                 D4   125    2
## 108 2831                    Taylor Creative    35    2
## 109 2877                      Myriad Supply    44    2
## 110 2896                               Telx   316    2
## 111 2924                        WTP Advisors    90    2
## 112 2981                   Flexible Systems    95    2
## 113 2997                  Fibertech Networks   250    2
## 114 3009                  5Linx Enterprises   268    2
## 115 3054                      DDS Companies   219    2
```

```
## 116 3097                            Gates Automotive    64    2
## 117 3103                             SmartSign.com       34    2
## 118 3117                             Team Epiphany       37    2
## 119 3153                      Juice Pharma Worldwide     186    2
## 120 3168                            Atrium Staffing     107    2
## 121 3210                             Dance With Me       50    2
## 122 3262                            Sriven Systems      210    2
## 123 3289                               Butler/Till       73    2
## 124 3351                         Bespoke Education       70    2
## 125 3355                          Home Clean Home       55    2
## 126 3358                                 HR Works       53    2
## 127 3383                          Graduationsource       55    2
## 128 3401                   Alphaserve Technologies       80    2
## 129 3424                       Stemp Systems Group       39    2
## 130 3535                             Tully Rinckey       87    2
## 131 3592                             Liberty Pumps      132    2
## 132 3594                                 Post Beam       32    2
## 133 3601                             Best Cleaners       85    2
## 134 3603                            Gerstein Fisher       35    2
## 135 3632            Integrated Systems Management      144    2
## 136 3643                                 Docutrend       45    2
## 137 3661 Environmental Products & Services of Vermont   250    2
## 138 3663                          Liquid Technology       33    2
## 139 3685               Eminent Technology Solutions       48    2
## 140 3764          American Portfolios Financial Services   81    2
## 141 3787                           Optimax Systems      200    2
## 142 3864                             ClassBook.com       61    2
## 143 3871                      eTransMedia Technology     298    2
## 144 3891                                 Net@Work      156    2
## 145 3924             Finger Lakes Technologies Group     105    2
## 146 3943                             Gersh Academy       51    2
## 147 4000                             Send Word Now       59    2
## 148 4007                      Spear Physical Therapy      40    2
## 149 4008                             Nutricap Labs       45    2
## 150 4117                                NorthPoint      144    2
## 151 4166                            Motivators.com       55    2
## 152 4170               Creative Environment Solutions     60    2
## 153 4245             Computer Resources of America       47    2
## 154 4287                                LinguaLinx       32    2
## 155 4335                         New York Kids Club      200    2
## 156 4336                          Cyrus Innovation       46    2
## 157 4352                                   Kitware      110    2
## 158 4399                             Direct Agents       47    2
## 159 4422                          BPA International      139    2
## 160 4448                         The Atlantic Group       51    2
## 161 4474                    Sam Schwartz Engineering      94    2
## 162 4490                       Catseye Pest Control       62    2
## 163 4544               ePromos Promotional Products      104    2
## 164 4552                                  Paradysz      270    2
## 165 4560           Sentry Communications & Security       40    2
## 166 4569                        Bogota Latin Bistro       63    2
## 167 4595                                   Numerix      198    2
## 168 4602                            Brite Computers       54    2
## 169 4605                              QED National       89    2
```

```
## 170 4611                                       Inergex    140    2
## 171 4632                                    Transfinder     63    2
## 172 4639  American Petroleum Equipment & Construction      68    2
## 173 4652                                    SaveAround      64    2
## 174 4732                                       Mycroft     275    2
## 175 4738                                   McMurry/TMG     269    2
## 176 4755                               Greyston Bakery      84    2
## 177 4774                               The Lloyd Group      73    2
## 178 4803                               Benefit Resource     96    2
## 179 4820                                  Aluf Plastics    307    2
## 180 4839                                      Makovsky      55    2
## 181 4849                            GTM Payroll Services    43    2
## 182 4891                     Express Air Freight Unlmited    70    2
## 183 4896                   Rand Engineering & Architecture   80    2
## 184 4934                                 Buffalo Filter     67    2
## 185 4944                                McElroy Deutsch    288    2
## 186 4981  SmartSource Computer & Audio Visual Rentals     296    2
## 187   26                                   BeenVerified     17    1
## 188   37                                   YellowHammer     27    1
## 189  124                               Quantum Networks     28    1
## 190  126                        Renegade Furniture Group    17    1
## 191  232                                 Rethink Autism     22    1
## 192  250                                  Gravity Media     15    1
## 193  264                                      MSR Promo      4    1
## 194  267                                     NUE Agency      4    1
## 195  280                                   RosettaBooks      7    1
## 196  299                     Ultraflex Power Technologies   12    1
## 197  309                                       Ivy Exec     30    1
## 198  388                        Odyssey Telecommunications    6    1
## 199  400                                 Forward Health      6    1
## 200  418                                          Grok     19    1
## 201  465                                  Bustin Boards     25    1
## 202  490                             MASS Communications     25    1
## 203  502                                      Largetail     12    1
## 204  515                               Integra Networks     31    1
## 205  526                                Surface Magazine     14    1
## 206  568                                  LocalResponse     25    1
## 207  709                                   Driven Local     28    1
## 208  714                                The Classy Home     18    1
## 209  728                              Vector Media Group     12    1
## 210  744                                    BulbAmerica     22    1
## 211  769                                  Peeled Snacks     14    1
## 212  788                                 Digital Energy     16    1
## 213  809                                 The Ruby Group     15    1
## 214  818                             Salzinger & Company     15    1
## 215  832                           In the Know Experiences    6    1
## 216  854                                      Hipercept     30    1
## 217  896                                      Optimatic      8    1
## 218  971                         Imagine Easy Solutions     19    1
## 219 1010                               Resolute Digital     28    1
## 220 1020                                   BarkerGilmore     12    1
## 221 1067                             Dionysus Consulting     22    1
## 222 1100                                    ArkNet Media     20    1
## 223 1166                         Manhattan Strategy Group    30    1
```

```
## 224 1181                            Sub Rosa     28     1
## 225 1203                            InRhythm     27     1
## 226 1210                            Trepoint     15     1
## 227 1214                    ALAC International     11     1
## 228 1234                            Berricle     11     1
## 229 1443                     Oliver Staffing      7     1
## 230 1458                   Dynomighty Design     12     1
## 231 1485                 Cipher Tech Solutions     17     1
## 232 1540                           Gaffos.com     15     1
## 233 1559                             ModiFace     25     1
## 234 1612                     Andiamo Partners      8     1
## 235 1622                      LJM Consultants     22     1
## 236 1646    New York Technology Partners - Rochester     23     1
## 237 1709                          Tribeca Nik      5     1
## 238 1717                       Society Awards     10     1
## 239 1723                         Spicer Group     20     1
## 240 1767                       Vcorp Services     15     1
## 241 1877                          MaeTec Power     22     1
## 242 1879                      Durante Rentals     25     1
## 243 1920             Reliance Star Payment Services     14     1
## 244 1954                           CleanEdison     25     1
## 245 2077                          Sachin Babi     23     1
## 246 2089                          React2Media     15     1
## 247 2234                           WorkCompEDI     13     1
## 248 2292                 IntraLogic Solutions     25     1
## 249 2352                      LDJ Productions     12     1
## 250 2432                     Bear Metal Works     13     1
## 251 2452                          Cafe Grumpy     30     1
## 252 2503              Marvin Traub Associates     12     1
## 253 2549             Central Park West Dentistry     20     1
## 254 2561                           1st Equity      7     1
## 255 2562         Advantage Communications Group     15     1
## 256 2585                       Primary Support     14     1
## 257 2622                            Leviathan     11     1
## 258 2656                             TekScape     26     1
## 259 2666            Adventure in Food Trading     12     1
## 260 2675                Cedar Petrochemcials      5     1
## 261 2702                        Ryan Printing     26     1
## 262 2706                  Primary Wave Media     15     1
## 263 2747                     Stage 2 Networks     20     1
## 264 2764                   Weiss & Associates      9     1
## 265 2784                         SportsSignup     21     1
## 266 2799            School Choice International     21     1
## 267 2889                   Logistic Dynamics     25     1
## 268 3056                           VHMnetwork     10     1
## 269 3060                            Nasty Pig     10     1
## 270 3115                       MyOTCStore.com     15     1
## 271 3146                     Verge Marketing     21     1
## 272 3193                 Access Display Group     22     1
## 273 3251                    Modus Associates      9     1
## 274 3313                   MyDailyMoment.com     16     1
## 275 3532                    The Urban Group     24     1
## 276 3542                           SweetRiot      5     1
## 277 3544                Proactive Technologies     24     1
```

```
## 278 3549                        Influence Graphics        24        1
## 279 3716                           LISS Consulting        18        1
## 280 3789                              Xtreme Trips        30        1
## 281 3925                                      Sinu        25        1
## 282 4054                                      BUMI         8        1
## 283 4128                       Cvision Technologies        28        1
## 284 4154      Business Communications Management        21        1
## 285 4158                                  Edulence        15        1
## 286 4162                                JR Products        30        1
## 287 4250                                   NiCoForm        16        1
## 288 4307                                ItsHot.com         9        1
## 289 4432                                C2C Outdoor        12        1
## 290 4447                        Select Office Suites        21        1
## 291 4467                             Blue Telescope         9        1
## 292 4471                   KJ Technology Consulting        13        1
## 293 4475                      Presentation Products        29        1
## 294 4522                           Geneva Worldwide        30        1
## 295 4559                                    Tabush        23        1
## 296 4565                                EquipSystems        31        1
## 297 4645                               DatapointLabs        11        1
## 298 4670                                   VarData        18        1
## 299 4671                                 Webucator        24        1
## 300 4702      Copier Fax Business Technologies        27        1
## 301 4728                        DurAmerica Brokerage        15        1
## 302 4731                                 PlayScripts        10        1
## 303 4782      Life Safety Engineered Systems        25        1
## 304 4800                    Documentation Strategies        25        1
## 305 4853                                     Affect        18        1
## 306 4950      Colorspec Coatings International        10        1
## 307 4954                         Premium Productions        16        1
## 308  632                             Armadale Capital         2        0
## 309 1028                    Direct Transport Systems         1        0
## 310 1095                         UpSurge Media Group         2        0
## 311 1986                        Medical Supply Depot         3        0
##      Growth_Rate   Revenue                       Industry          City State 2
## 1           0.48 5.976e+08 Business Products & Services      Pittsford    NY 2
## 2           0.36 4.600e+09 Consumer Products & Services       New York    NY 2
## 3           3.94 1.671e+08                   IT Services       Syracuse    NY 2
## 4           2.66 2.149e+08               Human Resources       New York    NY 2
## 5           2.38 9.040e+07                   IT Services       New York    NY 2
## 6           1.67 4.500e+06                      Software       New York    NY 2
## 7           1.22 2.430e+07 Business Products & Services       New York    NY 2
## 8           1.14 3.070e+07                      Software       New York    NY 2
## 9           1.14 5.270e+07            Financial Services       Fairport    NY 2
## 10          1.13 2.310e+08               Human Resources       New York    NY 2
## 11          1.11 4.590e+07                   IT Services       New York    NY 2
## 12          1.05 5.530e+07               Human Resources       New York    NY 2
## 13          0.92 1.453e+08                   IT Services       New York    NY 2
## 14          0.88 2.360e+07               Food & Beverage       New York    NY 2
## 15          0.84 6.790e+07                   IT Services       New York    NY 2
## 16          0.79 3.570e+07               Human Resources       New York    NY 2
## 17          0.71 2.808e+08          Travel & Hospitality       New York    NY 2
## 18          0.67 5.860e+07          Travel & Hospitality       New York    NY 2
## 19          0.66 3.496e+08 Business Products & Services       New York    NY 2
```

```
## 20        0.62 1.464e+08                        Media         New York  NY 2
## 21        0.60 3.465e+08 Business Products & Services         New York  NY 2
## 22        0.55 3.413e+08 Business Products & Services         New York  NY 2
## 23        0.52 3.043e+08                     Software         Uniondale  NY 2
## 24        0.49 1.516e+08          Financial Services         New York  NY 2
## 25        0.46 8.770e+07 Business Products & Services         New York  NY 2
## 26        0.44 3.800e+09                  IT Services        Tarrytown  NY 2
## 27        0.44 1.400e+07                     Security         New York  NY 2
## 28        0.43 2.571e+08        Travel & Hospitality         New York  NY 2
## 29        0.37 3.520e+08 Business Products & Services     White Plains  NY 2
## 30       73.22 8.100e+06        Advertising & Marketing         New York  NY 2
## 31       67.02 7.100e+06        Advertising & Marketing         New York  NY 2
## 32       53.65 5.900e+06          Financial Services        Rock Hill  NY 2
## 33       44.99 2.790e+07        Advertising & Marketing         New York  NY 2
## 34       44.85 6.900e+06        Advertising & Marketing         New York  NY 2
## 35       25.13 1.540e+07        Travel & Hospitality         Brooklyn  NY 2
## 36       23.01 1.660e+07                        Media         New York  NY 2
## 37       19.28 7.500e+06                  IT Services         Fishkill  NY 2
## 38       15.85 1.630e+07          Telecommunications           Walden  NY 2
## 39       15.28 2.370e+07        Advertising & Marketing         New York  NY 2
## 40       13.77 4.000e+06 Business Products & Services         New York  NY 2
## 41       10.24 7.200e+06        Advertising & Marketing         Brooklyn  NY 2
## 42        9.19 2.875e+08          Financial Services         New York  NY 2
## 43        8.70 6.250e+07 Consumer Products & Services         New York  NY 2
## 44        8.70 5.100e+06             Food & Beverage  Pleasant Valley  NY 2
## 45        7.54 3.880e+07                       Energy    Ballston Lake  NY 2
## 46        7.01 2.170e+07          Financial Services         New York  NY 2
## 47        6.98 3.670e+07          Financial Services         New York  NY 2
## 48        6.94 2.900e+06                       Health      Floral Park  NY 2
## 49        6.84 6.700e+06                        Media         New York  NY 2
## 50        6.52 5.250e+07                        Media         New York  NY 2
## 51        6.42 2.240e+07          Financial Services Long Island City  NY 2
## 52        6.13 1.890e+07                        Media         New York  NY 2
## 53        6.12 2.500e+06                    Education         New York  NY 2
## 54        5.55 9.500e+06        Advertising & Marketing Saratoga Springs  NY 2
## 55        5.40 6.600e+06        Advertising & Marketing         New York  NY 2
## 56        5.04 1.700e+07                  IT Services           Ithaca  NY 2
## 57        5.00 4.040e+07          Financial Services         New YOrk  NY 2
## 58        4.75 5.890e+07                       Retail         New York  NY 2
## 59        4.67 5.400e+06                       Energy         Oneonta  NY 2
## 60        4.11 7.000e+06                  IT Services         New York  NY 2
## 61        4.06 5.100e+06                        Media         New York  NY 2
## 62        4.03 5.240e+07 Consumer Products & Services         new york  NY 2
## 63        3.85 1.420e+07        Advertising & Marketing         New York  NY 2
## 64        3.75 8.700e+06                Manufacturing         New York  NY 2
## 65        3.72 5.000e+06        Advertising & Marketing         New York  NY 2
## 66        3.52 5.900e+07                  IT Services         New York  NY 2
## 67        3.25 1.380e+07                       Retail         New York  NY 2
## 68        3.14 1.490e+07        Advertising & Marketing         New York  NY 2
## 69        2.94 1.290e+07                     Software         New York  NY 2
## 70        2.92 2.000e+06                    Education         New York  NY 2
## 71        2.82 3.670e+07        Advertising & Marketing         New York  NY 2
## 72        2.78 2.810e+07                    Insurance         St James  NY 2
## 73        2.76 6.300e+06                       Health     White Plains  NY 2
```

```
## 74     2.65 6.800e+06       Advertising & Marketing      IRVINGTON   NY 2
## 75     2.54 6.730e+07       Advertising & Marketing       New York   NY 2
## 76     2.47 6.500e+06       Advertising & Marketing       New York   NY 2
## 77     2.27 2.090e+07       Advertising & Marketing       New York   NY 2
## 78     2.19 5.600e+06                    IT Services         Latham   NY 2
## 79     2.03 5.500e+06                Food & Beverage      Canastota   NY 2
## 80     2.00 5.320e+07                    IT Services   Clifton Park   NY 2
## 81     1.89 5.100e+06                      Education       New York   NY 2
## 82     1.89 1.408e+08                         Energy      Getzville   NY 2
## 83     1.81 8.100e+07                  Manufacturing       New York   NY 2
## 84     1.76 3.800e+06             Financial Services       New York   NY 2
## 85     1.73 8.900e+06           Travel & Hospitality     Ronkonkoma   NY 2
## 86     1.70 1.410e+07                         Health   Lake Success   NY 2
## 87     1.69 1.650e+07                         Retail    Cheektowaga   NY 2
## 88     1.68 7.200e+06 Business Products & Services    Honeoye Falls   NY 2
## 89     1.67 2.550e+07       Advertising & Marketing       New York   NY 2
## 90     1.60 1.820e+07                       Software  Staten Island   NY 2
## 91     1.56 3.080e+07                Human Resources       New York   NY 2
## 92     1.52 4.700e+06       Advertising & Marketing       New York   NY 2
## 93     1.51 1.070e+07       Advertising & Marketing       New York   NY 2
## 94     1.50 9.900e+06       Advertising & Marketing       New York   NY 2
## 95     1.48 1.200e+07                    IT Services       New York   NY 2
## 96     1.45 5.200e+07                    IT Services      Rochester   NY 2
## 97     1.42 1.880e+07 Business Products & Services        New York   NY 2
## 98     1.42 1.470e+07                    IT Services      Scarsdale   NY 2
## 99     1.40 1.410e+07             Telecommunications       Fairport   NY 2
## 100    1.39 6.200e+06                         Health          Utica   NY 2
## 101    1.39 3.120e+07                         Energy    New Windsor   NY 2
## 102    1.31 2.760e+07                         Health           troy   NY 2
## 103    1.29 9.700e+06       Advertising & Marketing       New York   NY 2
## 104    1.27 1.350e+07       Advertising & Marketing       New York   NY 2
## 105    1.24 1.122e+08                         Health       New York   NY 2
## 106    1.23 3.290e+07       Advertising & Marketing       New York   NY 2
## 107    1.22 1.430e+07                    IT Services      Rochester   NY 2
## 108    1.22 9.200e+06       Advertising & Marketing       New York   NY 2
## 109    1.19 2.290e+07              Computer Hardware       New York   NY 2
## 110    1.18 2.143e+08             Telecommunications       New York   NY 2
## 111    1.16 1.600e+07 Business Products & Services    White Plains   NY 2
## 112    1.13 2.960e+07                    IT Services      Hauppauge   NY 2
## 113    1.13 1.423e+08             Telecommunications      ROCHESTER   NY 2
## 114    1.12 1.036e+08             Telecommunications      Rochester   NY 2
## 115    1.09 5.250e+07                   Construction      Rochester   NY 2
## 116    1.07 6.900e+06 Consumer Products & Services      Henrietta   NY 2
## 117    1.07 2.530e+07 Business Products & Services       Brooklyn   NY 2
## 118    1.06 8.200e+06       Advertising & Marketing       New York   NY 2
## 119    1.04 5.380e+07       Advertising & Marketing       New York   NY 2
## 120    1.03 9.700e+07                Human Resources       New York   NY 2
## 121    1.01 4.000e+06                      Education       New York   NY 2
## 122    0.98 2.530e+07                    IT Services       Melville   NY 2
## 123    0.97 1.236e+08       Advertising & Marketing      Rochester   NY 2
## 124    0.94 2.600e+06                      Education       New York   NY 2
## 125    0.94 2.100e+06 Consumer Products & Services       Brooklyn   NY 2
## 126    0.93 4.300e+06                Human Resources       Fairport   NY 2
## 127    0.92 5.300e+06                      Education   Port Chester   NY 2
```

```
## 128              0.92 1.640e+07                 IT Services         New York   NY 2
## 129             19.37 6.800e+06                 IT Services Long Island City   NY 2
## 130              0.86 8.800e+06 Consumer Products & Services         Albany   NY 2
## 131              0.84 6.080e+07               Manufacturing         Bergen   NY 2
## 132              0.84 5.900e+06       Advertising & Marketing         New York   NY 2
## 133              0.84 4.800e+06 Consumer Products & Services     Schenectady   NY 2
## 134              0.84 9.700e+06           Financial Services         New York   NY 2
## 135              0.83 4.600e+06                 IT Services       Tarrytown   NY 2
## 136              0.82 8.200e+06 Business Products & Services         New York   NY 2
## 137              0.81 4.510e+07       Environmental Services       Syracuse   NY 2
## 138              0.81 1.000e+07                 IT Services        Brooklyn   NY 2
## 139              0.80 6.300e+06                 IT Services         Buffalo   NY 2
## 140              0.77 1.170e+08           Financial Services        Holbrook   NY 2
## 141              0.76 2.010e+07               Manufacturing         Ontario   NY 2
## 142              0.73 9.600e+06                   Education       Castleton   NY 2
## 143              0.72 2.590e+07                      Health           Troy   NY 2
## 144              0.72 2.930e+07                 IT Services         New York   NY 2
## 145              0.70 2.310e+07           Telecommunications         Victor   NY 2
## 146              0.70 8.600e+06                   Education      Huntington   NY 2
## 147              0.67 1.660e+07           Telecommunications         New York   NY 2
## 148              0.67 5.200e+06                      Health         New York   NY 2
## 149              0.67 3.000e+07                      Health     Farmingdale   NY 2
## 150              0.64 2.730e+07 Business Products & Services         New York   NY 2
## 151              0.62 1.320e+07       Advertising & Marketing        westbury   NY 2
## 152              0.62 5.300e+06       Environmental Services         New York   NY 2
## 153              0.59 7.700e+06                 IT Services         New York   NY 2
## 154              0.57 4.300e+06 Business Products & Services         Cohoes   NY 2
## 155              0.56 1.250e+07                   Education         New York   NY 2
## 156              0.56 5.900e+06                   Software        new york   NY 2
## 157              0.55 2.130e+07                   Software    Clifton Park   NY 2
## 158              0.54 2.270e+07       Advertising & Marketing         New York   NY 2
## 159              0.53 5.400e+06           Telecommunications     Carle Place   NY 2
## 160              0.52 6.430e+07 Business Products & Services         New York   NY 2
## 161              0.51 1.680e+07                 Engineering         New York   NY 2
## 162              0.51 9.000e+06 Consumer Products & Services       Castleton   NY 2
## 163              0.49 2.840e+07       Advertising & Marketing         New York   NY 2
## 164              0.49 4.100e+07       Advertising & Marketing         New York   NY 2
## 165              0.48 6.000e+06                    Security      Hicksville   NY 2
## 166              0.48 3.800e+06             Food & Beverage        Brooklyn   NY 2
## 167              0.48 5.640e+07                   Software         New York   NY 2
## 168              0.47 1.990e+07                 IT Services         Victor   NY 2
## 169              0.47 6.700e+06                 IT Services         New York   NY 2
## 170              0.47 1.740e+07                 IT Services         Buffalo   NY 2
## 171              0.47 7.800e+06                   Software     Schenectady   NY 2
## 172              0.46 1.040e+07                Construction         Walden   NY 2
## 173              0.46 6.900e+06       Advertising & Marketing      Binghamton   NY 2
## 174              0.43 2.640e+07                 IT Services         New York   NY 2
## 175              0.43 9.140e+07       Advertising & Marketing   New York City   NY 2
## 176              0.42 1.010e+07             Food & Beverage         Yonkers   NY 2
## 177              0.42 9.900e+06                 IT Services         New York   NY 2
## 178              0.41 1.330e+07 Business Products & Services       Rochester   NY 2
## 179              0.40 1.029e+08               Manufacturing      Orangeburg   NY 2
## 180              0.40 1.430e+07       Advertising & Marketing         New York   NY 2
## 181              0.40 6.700e+06           Financial Services    Clifton Park   NY 2
```

```
## 182      0.38 2.860e+07      Logistics & Transportation       Jamaica   NY 2
## 183      0.38 9.800e+06                    Engineering       New York   NY 2
## 184      0.37 1.500e+07                  Manufacturing      Lancaster   NY 2
## 185      0.36 1.165e+08 Business Products & Services       New York   NY 2
## 186      0.35 5.730e+07 Business Products & Services      Hauppauge   NY 2
## 187     84.43 1.370e+07 Consumer Products & Services       New York   NY 2
## 188     67.40 1.800e+07      Advertising & Marketing       New York   NY 2
## 189     29.44 1.150e+07            Telecommunications       New York   NY 2
## 190     29.26 9.800e+06                        Retail        Hewlett   NY 2
## 191     18.27 4.100e+06                     Education       New York   NY 2
## 192     17.24 1.240e+07      Advertising & Marketing       New York   NY 2
## 193     16.30 2.400e+06 Business Products & Services       New York   NY 2
## 194     16.04 4.500e+06                         Media       New York   NY 2
## 195     15.49 3.200e+06                         Media       New York   NY 2
## 196     14.54 2.200e+06                 Manufacturing     Ronkonkoma   NY 2
## 197     14.01 2.200e+06               Human Resources       New York   NY 2
## 198     11.40 3.300e+06            Telecommunications         colden   NY 2
## 199     11.07 1.050e+07                        Health       New York   NY 2
## 200     10.77 9.400e+06      Advertising & Marketing       New York   NY 2
## 201      9.92 4.000e+06 Consumer Products & Services       Brooklyn   NY 2
## 202      9.33 9.800e+06            Telecommunications       New York   NY 2
## 203      9.10 3.700e+06      Advertising & Marketing       New York   NY 2
## 204      8.83 1.660e+07            Telecommunications         Latham   NY 2
## 205      8.63 3.900e+06                         Media       New York   NY 2
## 206      7.94 3.900e+06      Advertising & Marketing       new york   NY 2
## 207      6.45 7.500e+06      Advertising & Marketing       Melville   NY 2
## 208      6.43 2.800e+06 Consumer Products & Services       Brooklyn   NY 2
## 209      6.30 2.100e+06      Advertising & Marketing       New York   NY 2
## 210      6.16 6.300e+06                        Retail       Brooklyn   NY 2
## 211      5.94 5.800e+06               Food & Beverage       Brooklyn   NY 2
## 212      5.79 1.230e+07 Consumer Products & Services       Brooklyn   NY 2
## 213      5.57 1.100e+07                   Real Estate         Goshen   NY 2
## 214      5.53 3.900e+06                   IT Services      Scarsdale   NY 2
## 215      5.39 4.300e+06          Travel & Hospitality       New York   NY 2
## 216      5.23 4.200e+06                   Real Estate       New York   NY 2
## 217      4.97 8.700e+06      Advertising & Marketing       new york   NY 2
## 218      4.45 3.600e+06                     Education  New York City   NY 2
## 219      4.26 7.100e+06      Advertising & Marketing       New York   NY 2
## 220      4.22 2.100e+06               Human Resources       Fairport   NY 2
## 221      3.96 3.400e+06      Advertising & Marketing     West Islip   NY 2
## 222      3.82 5.300e+06 Consumer Products & Services    Garden City   NY 2
## 223      3.60 6.300e+06 Business Products & Services       New York   NY 2
## 224      3.56 9.500e+06      Advertising & Marketing       New York   NY 2
## 225      3.47 3.200e+06                   IT Services       New York   NY 2
## 226      3.43 4.300e+06      Advertising & Marketing     Great Neck   NY 2
## 227      3.42 5.650e+07                 Manufacturing       New York   NY 2
## 228      3.34 2.200e+06                        Retail        ASTORIA   NY 2
## 229      2.78 4.600e+06               Human Resources       New York   NY 2
## 230      2.74 4.000e+06                        Retail       Brooklyn   NY 2
## 231      2.69 2.700e+06           Government Services          Nyack   NY 2
## 232      2.58 4.400e+06                        Retail         Inwood   NY 2
## 233      2.55 2.100e+06                      Software       New York   NY 2
## 234      2.46 6.200e+06               Human Resources       New York   NY 2
## 235      2.43 4.200e+06      Logistics & Transportation      Melville   NY 2
```

```
## 236      2.37 3.200e+06                 IT Services         Rochester NY 2
## 237      2.27 4.700e+06 Consumer Products & Services       Lindenhurst NY 2
## 238      2.26 3.300e+06 Consumer Products & Services Long Island City NY 2
## 239      2.25 5.600e+06                Construction           Buffalo NY 2
## 240      2.17 6.300e+06 Business Products & Services           Monsey NY 2
## 241      2.03 1.010e+07           Telecommunications          Fairport NY 2
## 242      2.03 4.800e+06                Construction             Bronx NY 2
## 243      1.97 2.000e+06          Financial Services   Oakland Gardens NY 2
## 244      1.94 3.500e+06                   Education          New York NY 2
## 245      1.80 4.700e+06 Consumer Products & Services          New York NY 2
## 246      1.78 5.600e+06       Advertising & Marketing         New York NY 2
## 247      1.66 2.900e+06                 IT Services          Holbrook NY 2
## 248      1.59 3.700e+06                    Security         Massapequa NY 2
## 249      1.54 1.230e+07 Business Products & Services          New York NY 2
## 250      1.48 2.400e+06               Manufacturing           Buffalo NY 2
## 251      1.46 2.200e+06             Food & Beverage          Brooklyn NY 2
## 252      1.42 3.800e+06                      Retail          New York NY 2
## 253      1.39 3.100e+06                      Health          New York NY 2
## 254      1.38 2.100e+06                 Real Estate           Syosset NY 2
## 255      1.38 5.600e+06          Telecommunications    Roslyn Heights NY 2
## 256      1.36 2.300e+06                 IT Services          New York NY 2
## 257      1.34 4.840e+07 Business Products & Services          Brooklyn NY 2
## 258      1.32 1.360e+07                 IT Services          New York NY 2
## 259      1.32 2.600e+06             Food & Beverage           Menands NY 2
## 260      1.31 2.037e+08                      Energy          New York NY 2
## 261      1.30 7.300e+06               Manufacturing          blauvelt NY 2
## 262      1.30 3.000e+06       Advertising & Marketing      Pleasantville NY 2
## 263      1.27 9.200e+06          Telecommunications          New York NY 2
## 264      1.26 2.000e+06 Consumer Products & Services          New York NY 2
## 265      1.25 3.300e+06                    Software Saratoga Springs NY 2
## 266      1.24 2.800e+06                   Education      White Plains NY 2
## 267      1.18 3.960e+07    Logistics & Transportation           Amherst NY 2
## 268      1.09 4.500e+06       Advertising & Marketing         New York NY 2
## 269      1.09 2.000e+06                      Retail          New York NY 2
## 270      1.06 3.140e+07                      Retail          New York NY 2
## 271      1.04 4.000e+06       Advertising & Marketing         New York NY 2
## 272      1.02 4.100e+06               Manufacturing          Freeport NY 2
## 273      0.99 2.300e+06       Advertising & Marketing         New York NY 2
## 274      0.96 5.900e+06                       Media          New York NY 2
## 275      0.86 5.700e+06                Construction         Woodhaven NY 2
## 276      0.86 2.000e+06             Food & Beverage          New York NY 2
## 277      0.86 4.000e+06                 IT Services          New York NY 2
## 278      0.86 4.100e+06       Advertising & Marketing         New York NY 2
## 279      0.79 2.400e+06                 IT Services      Lake Success NY 2
## 280      0.76 6.700e+06        Travel & Hospitality          plainview NY 2
## 281      0.70 3.000e+06                 IT Services          New York NY 2
## 282      0.65 2.300e+06                 IT Services          New York NY 2
## 283      0.63 5.200e+06                    Software      Forest Hills NY 2
## 284      0.62 1.090e+07          Telecommunications          New York NY 2
## 285      0.62 2.000e+06                    Software          New York NY 2
## 286      0.62 4.900e+06               Manufacturing   Clarence Center NY 2
## 287      0.59 2.100e+06               Manufacturing         Rochester NY 2
## 288      0.57 5.700e+06                      Retail          New York NY 2
## 289      0.53 2.770e+07       Advertising & Marketing         New York NY 2
```

```
## 290       0.52 7.600e+06                  Real Estate    New York   NY 2
## 291       0.52 3.400e+06                        Media    New York   NY 2
## 292       0.51 4.000e+06                  IT Services    New York   NY 2
## 293       0.51 9.700e+06                  Engineering    New York   NY 2
## 294       0.50 7.400e+06 Business Products & Services    new york   NY 2
## 295       0.48 3.900e+06                  IT Services    New York   NY 2
## 296       0.48 3.000e+06                       Health    New York   NY 2
## 297       0.46 2.500e+06                  Engineering      Ithaca   NY 2
## 298       0.45 1.480e+07            Telecommunications   Rochester   NY 2
## 299       0.45 4.600e+06                    Education  Jamesville   NY 2
## 300       0.44 5.300e+06 Business Products & Services      Buffalo   NY 2
## 301       0.43 2.700e+06                    Insurance Long Island City   NY 2
## 302       0.43 2.600e+06                       Retail    New York   NY 2
## 303       0.42 7.100e+06                     Security      buffalo   NY 2
## 304       0.41 2.300e+06                  IT Services   Rensselaer   NY 2
## 305       0.39 2.600e+06      Advertising & Marketing    New York   NY 2
## 306       0.36 3.300e+06                 Construction    Holbrook   NY 2
## 307       0.36 6.800e+06      Advertising & Marketing Pleasantville   NY 2
## 308       7.27 3.600e+06                       Health    New York   NY 2
## 309       4.15 2.800e+06   Logistics & Transportation  Farmingdale   NY 2
## 310       3.84 3.600e+06      Advertising & Marketing    New York   NY 2
## 311       1.89 2.600e+06                       Retail     Brooklyn   NY 2
```

Apply customer `groupby_column()` function

```
(inc_ny_emp_log_grp <- groupby_column(inc_ny_emp_log, Industry, emp_log))
```

```
## # A tibble: 25 x 6
##    Industry                    n_grp min_grp max_grp mean_grp med_grp
##    <chr>                       <int>   <dbl>   <dbl>    <dbl>   <dbl>
##  1 Business Products & Services   26       1       5     2.08       2
##  2 Consumer Products & Services   17       1       4     1.59       1
##  3 Financial Services             13       1       3     2.08       2
##  4 Food & Beverage                 9       1       3     1.67       2
##  5 Human Resources                11       1       3     2          2
##  6 IT Services                    43       1       3     1.84       2
##  7 Media                          11       1       3     1.64       2
##  8 Security                        4       1       3     1.75     1.5
##  9 Software                       13       1       3     1.92       2
## 10 Travel & Hospitality            7       1       3     2.14       2
## # ... with 15 more rows
```

Let's re-plot with custom `crossbar_plot()` function.

```
# plot by Median Employee by Industry
crossbar_plot(inc_ny_emp_log_grp,
              Industry, med_grp, med_grp, min_grp, max_grp,
              title = "Median Employee by Industry")
```

Median Employee by Industry
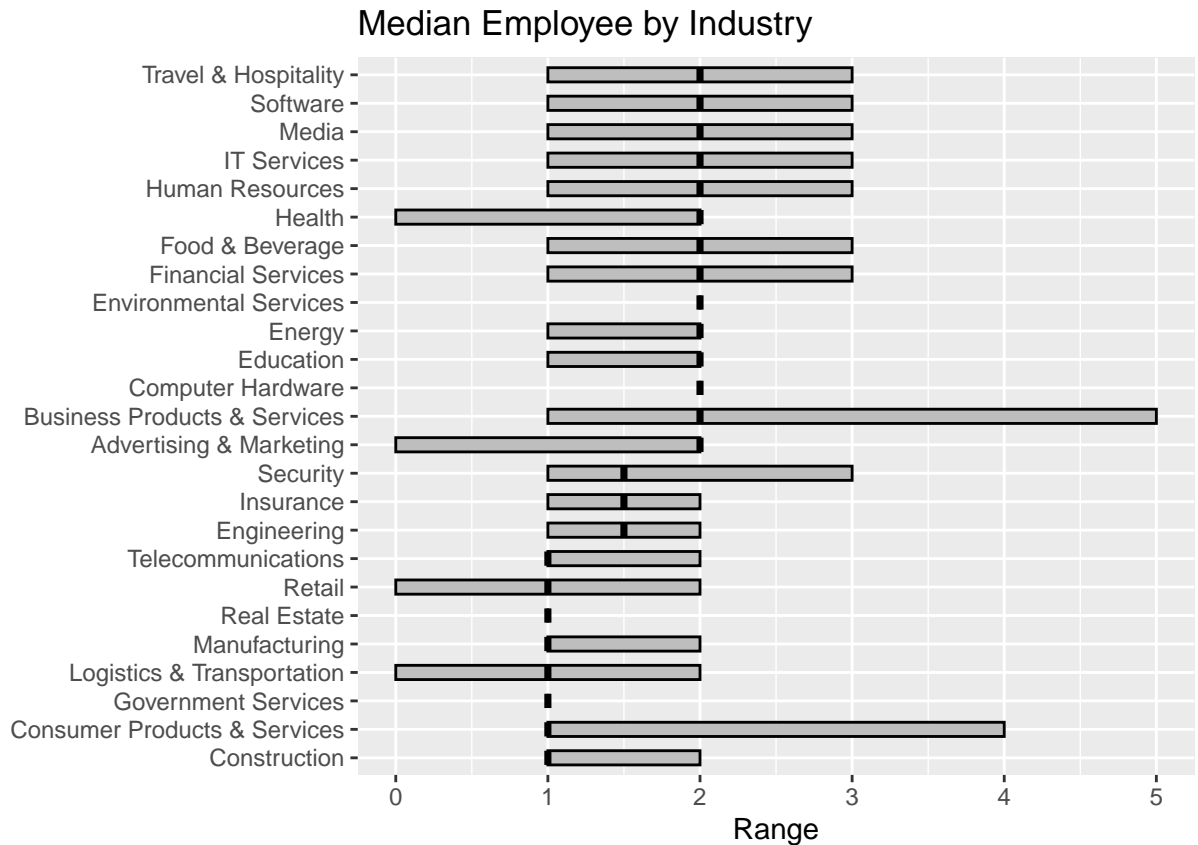
This is too generalized, also difficult to communicate range and median in log format to stakeholders.

Let's try dropping outliers

Drop Outliers

Let's classify outliers in the typical manner +- 1.5 * interquartile range

```
# find quartiles
quartiles <- stats::quantile(inc_ny$Employees, probs=c(.01,.99), na.rm=F)

# get IQR
IQR <- stats::IQR(inc_ny$Employees)

# define lower and upper outliers
lower <- quartiles[1] - 1.5*IQR
upper <- quartiles[2] + 1.5*IQR

# subset data set removing outliers
(inc_ny_no_outlier <-
  subset(inc_ny, inc_ny$Employees > lower & inc_ny$Employees < upper))
```

```
##       Rank                          Name Growth_Rate    Revenue
## 26      26                  BeenVerified       84.43   13700000
## 30      30                      Sailthru       73.22    8100000
## 37      37                   YellowHammer       67.40   18000000
```

```
## 38   38                       Conductor  67.02     7100000
## 48   48      Cinium Financial Services  53.65     5900000
## 70   70                         33Across  44.99    27900000
## 71   71                       LiveIntent  44.85     6900000
## 124  124              Quantum Networks  29.44    11500000
## 126  126       Renegade Furniture Group  29.26     9800000
## 153  153                     Regal Wings  25.13    15400000
## 174  174                      Refinery29  23.01    16600000
## 218  219             Novisync Solutions  19.28     7500000
## 231  232                 Rethink Autism  18.27     4100000
## 249  250                   Gravity Media  17.24    12400000
## 263  264                      MSR Promo  16.30     2400000
## 266  267                     NUE Agency  16.04     4500000
## 274  275            Technical Solutions  15.85    16300000
## 279  280                   RosettaBooks  15.49     3200000
## 284  285                            SET  15.28    23700000
## 298  299   Ultraflex Power Technologies  14.54     2200000
## 308  309                       Ivy Exec  14.01     2200000
## 313  314                        Panjiva  13.77     4000000
## 387  388   Odyssey Telecommunications  11.40     3300000
## 399  400                 Forward Health  11.07    10500000
## 417  418                           Grok  10.77     9400000
## 449  450                 Carrot Creative  10.24     7200000
## 464  465                  Bustin Boards   9.92     4000000
## 489  490            MASS Communications   9.33     9800000
## 496  497                    RCS Capital   9.19   287500000
## 501  502                       Largetail   9.10     3700000
## 514  515                Integra Networks   8.83    16600000
## 520  521           Happy Family Brands   8.70    62500000
## 521  522           Rockhedge Herb Farms   8.70     5100000
## 525  526               Surface Magazine   8.63     3900000
## 567  568                  LocalResponse   7.94     3900000
## 608  609               SmartWatt Energy   7.54    38800000
## 631  632               Armadale Capital   7.27     3600000
## 655  656             ForwardThink Group   7.01    21700000
## 659  660                         OnDeck   6.98    36700000
## 660  661            Quality In Real Time   6.94     2900000
## 668  669                     Touchstorm   6.84     6700000
## 700  701          Thrillist Media Group   6.52    52500000
## 708  709                   Driven Local   6.45     7500000
## 713  714                The Classy Home   6.43     2800000
## 716  717               Merchant Industry   6.42    22400000
## 727  728             Vector Media Group   6.30     2100000
## 743  744                     BulbAmerica   6.16     6300000
## 749  750                     Livestream   6.13    18900000
## 751  752                   Private Prep   6.12     2500000
## 768  769                   Peeled Snacks   5.94     5800000
## 787  788                 Digital Energy   5.79    12300000
## 808  809                 The Ruby Group   5.57    11000000
## 814  815                     Fingerpaint   5.55     9500000
## 817  818           Salzinger & Company   5.53     3900000
## 829  830          Launchpad Advertising   5.40     6600000
## 831  832       In the Know Experiences   5.39     4300000
## 853  854                       Hipercept   5.23     4200000
```

```
## 882   883                     Envisage Information Systems   5.04  17000000
## 889   890                                       Payoneer   5.00  40400000
## 895   896                                      Optimatic   4.97   8700000
## 921   922                                 Rebecca Minkoff   4.75  58900000
## 934   935                                          Ioxus   4.67   5400000
## 970   971                         Imagine Easy Solutions   4.45   3600000
## 1009  1010                              Resolute Digital   4.26   7100000
## 1019  1020                                  BarkerGilmore   4.22   2100000
## 1027  1028                       Direct Transport Systems   4.15   2800000
## 1035  1036                                         Enstoa   4.11   7000000
## 1047  1048                                   Curbed Media   4.06   5100000
## 1053  1054                                          Tarte   4.03  52400000
## 1066  1067                             Dionysus Consulting   3.96   3400000
## 1068  1069                            Systems Made Simple   3.94 167100000
## 1090  1091                            Blue Fountain Media   3.85  14200000
## 1094  1095                             UpSurge Media Group   3.84   3600000
## 1099  1100                                    ArkNet Media   3.82   5300000
## 1118  1119                                        Adafruit   3.75   8700000
## 1128  1129                                   Likeable Media   3.72   5000000
## 1165  1166                       Manhattan Strategy Group   3.60   6300000
## 1180  1181                                        Sub Rosa   3.56   9500000
## 1189  1190                                       Usablenet   3.52  59000000
## 1202  1203                                         InRhythm   3.47   3200000
## 1209  1210                                         Trepoint   3.43   4300000
## 1213  1214                             ALAC International   3.42  56500000
## 1233  1234                                         Berricle   3.34   2200000
## 1258  1259                                 Decorplanet.com   3.25  13800000
## 1292  1293                                    Madison Logic   3.14  14900000
## 1370  1371                                         Impelsys   2.94  12900000
## 1381  1382                             City Sounds of NY   2.92   2000000
## 1422  1423                                    Media6Degrees   2.82  36700000
## 1441  1442                           Global Warranty Group   2.78  28100000
## 1442  1443                                 Oliver Staffing   2.78   4600000
## 1450  1451                        Professional Data Systems   2.76   6300000
## 1457  1458                               Dynomighty Design   2.74   4000000
## 1484  1485                             Cipher Tech Solutions   2.69   2700000
## 1498  1499                             Sterling Infosystems   2.66 214900000
## 1502  1503                               AdCorp Media Group   2.65   6800000
## 1539  1540                                       Gaffos.com   2.58   4400000
## 1558  1559                                         ModiFace   2.55   2100000
## 1564  1565                                           Droga5   2.54  67300000
## 1608  1609                                 Enter:Marketing   2.47   6500000
## 1611  1612                                 Andiamo Partners   2.46   6200000
## 1621  1622                                   LJM Consultants   2.43   4200000
## 1639  1640                                          BlueWolf   2.38  90400000
## 1645  1646  New York Technology Partners - Rochester   2.37   3200000
## 1705  1706                                              MKG   2.27  20900000
## 1708  1709                                       Tribeca Nik   2.27   4700000
## 1716  1717                                   Society Awards   2.26   3300000
## 1722  1723                                       Spicer Group   2.25   5600000
## 1754  1755                       Deep Blue Communications   2.19   5600000
## 1766  1767                                    Vcorp Services   2.17   6300000
## 1875  1876                                        Servomation   2.03   5500000
## 1876  1877                                      MaeTec Power   2.03  10100000
```

```
## 1878 1879                          Durante Rentals  2.03   4800000
## 1894 1895                            nfrastructure  2.00  53200000
## 1919 1920            Reliance Star Payment Services  1.97   2000000
## 1953 1954                               CleanEdison  1.94   3500000
## 1981 1982                          Imagine Swimming  1.89   5100000
## 1984 1985                   U.S. Energy Development  1.89 140800000
## 1985 1986                       Medical Supply Depot  1.89   2600000
## 2066 2067                   Signature Systems Group  1.81  81000000
## 2076 2077                               Sachin Babi  1.80   4700000
## 2088 2089                                React2Media  1.78   5600000
## 2115 2116         Charter School Business Management  1.76   3800000
## 2141 2142                    Liberty Jet Management  1.73   8900000
## 2172 2173                              PM Pediatrics  1.70  14100000
## 2178 2179                         Christmas Central  1.69  16500000
## 2197 2198                                 KJT Group  1.68   7200000
## 2208 2209                                      inVNT  1.67  25500000
## 2217 2218                              Globo Mobile  1.67   4500000
## 2233 2234                               WorkCompEDI  1.66   2900000
## 2279 2280                        Celerant Technology  1.60  18200000
## 2291 2292                       IntraLogic Solutions  1.59   3700000
## 2335 2336                                BountyJobs  1.56  30800000
## 2351 2352                           LDJ Productions  1.54  12300000
## 2371 2372                   DiGennaro Communications  1.52   4700000
## 2398 2399                                      Rain  1.51  10700000
## 2404 2405                                  Sullivan  1.50   9900000
## 2431 2432                          Bear Metal Works  1.48   2400000
## 2435 2436                         Shinetech Software  1.48  12000000
## 2451 2452                               Cafe Grumpy  1.46   2200000
## 2474 2475                 Systems Management/Planning  1.45  52000000
## 2493 2494                                       RVM  1.42  18800000
## 2502 2503                    Marvin Traub Associates  1.42   3800000
## 2505 2506                                    QueBIT  1.42  14700000
## 2525 2526                       Soleo Communications  1.40  14100000
## 2541 2542                           The Fortus Group  1.39   6200000
## 2548 2549                  Central Park West Dentistry  1.39   3100000
## 2555 2556                 Precision Pipeline Solutions  1.39  31200000
## 2560 2561                                1st Equity  1.38   2100000
## 2561 2562           Advantage Communications Group  1.38   5600000
## 2584 2585                           Primary Support  1.36   2300000
## 2621 2622                                 Leviathan  1.34  48400000
## 2655 2656                                   TekScape  1.32  13600000
## 2665 2666                  Adventure in Food Trading  1.32   2600000
## 2674 2675                        Cedar Petrochemcials  1.31 203700000
## 2677 2678                            Treo Solutions  1.31  27600000
## 2701 2702                             Ryan Printing  1.30   7300000
## 2705 2706                        Primary Wave Media  1.30   3000000
## 2717 2718                                CreativeFeed  1.29   9700000
## 2746 2747                          Stage 2 Networks  1.27   9200000
## 2750 2751                                    IOMedia  1.27  13500000
## 2763 2764                         Weiss & Associates  1.26   2000000
## 2783 2784                                SportsSignup  1.25   3300000
## 2794 2795                                    Onco360  1.24 112200000
## 2798 2799                 School Choice International  1.24   2800000
## 2804 2805                    LeadDog Marketing Group  1.23  32900000
```

```
## 2827 2828                                                 D4   1.22  14300000
## 2829 2830                                         ReSource Pro   1.22  24300000
## 2830 2831                                       Taylor Creative   1.22   9200000
## 2876 2877                                         Myriad Supply   1.19  22900000
## 2888 2889                                      Logistic Dynamics   1.18  39600000
## 2895 2896                                                 Telx   1.18 214300000
## 2923 2924                                          WTP Advisors   1.16  16000000
## 2963 2964                                              DataArt   1.14  30700000
## 2970 2971                                              ConServe   1.14  52700000
## 2980 2981                                      Flexible Systems   1.13  29600000
## 2994 2995                                     Pride Technologies   1.13 231000000
## 2996 2997                                    Fibertech Networks   1.13 142300000
## 3008 3009                                     5Linx Enterprises   1.12 103600000
## 3020 3021                                                Linium   1.11  45900000
## 3053 3054                                         DDS Companies   1.09  52500000
## 3055 3056                                            VHMnetwork   1.09   4500000
## 3059 3060                                              Nasty Pig   1.09   2000000
## 3096 3097                                      Gates Automotive   1.07   6900000
## 3102 3103                                          SmartSign.com   1.07  25300000
## 3114 3115                                         MyOTCStore.com   1.06  31400000
## 3116 3117                                          Team Epiphany   1.06   8200000
## 3135 3136                                FSO Onsite Outsourcing   1.05  55300000
## 3145 3146                                        Verge Marketing   1.04   4000000
## 3152 3153                                  Juice Pharma Worldwide   1.04  53800000
## 3167 3168                                       Atrium Staffing   1.03  97000000
## 3192 3193                                    Access Display Group   1.02   4100000
## 3209 3210                                          Dance With Me   1.01   4000000
## 3250 3251                                        Modus Associates   0.99   2300000
## 3261 3262                                         Sriven Systems   0.98  25300000
## 3288 3289                                            Butler/Till   0.97 123600000
## 3312 3313                                      MyDailyMoment.com   0.96   5900000
## 3350 3351                                      Bespoke Education   0.94   2600000
## 3354 3355                                        Home Clean Home   0.94   2100000
## 3357 3358                                              HR Works   0.93   4300000
## 3382 3383                                        Graduationsource   0.92   5300000
## 3386 3387                                         Mitchell/Martin   0.92 145300000
## 3400 3401                                 Alphaserve Technologies   0.92  16400000
## 3423 3424                                     Stemp Systems Group  19.37   6800000
## 3500 3500                                        Magnolia Bakery   0.88  23600000
## 3532 3532                                        The Urban Group   0.86   5700000
## 3535 3535                                           Tully Rinckey   0.86   8800000
## 3542 3542                                              SweetRiot   0.86   2000000
## 3544 3544                                  Proactive Technologies   0.86   4000000
## 3549 3549                                      Influence Graphics   0.86   4100000
## 3584 3584                                                Infusion   0.84  67900000
## 3592 3592                                          Liberty Pumps   0.84  60800000
## 3594 3594                                              Post Beam   0.84   5900000
## 3601 3601                                          Best Cleaners   0.84   4800000
## 3603 3603                                        Gerstein Fisher   0.84   9700000
## 3632 3632                            Integrated Systems Management   0.83   4600000
## 3643 3643                                              Docutrend   0.82   8200000
## 3661 3661 Environmental Products & Services of Vermont   0.81  45100000
## 3663 3663                                       Liquid Technology   0.81  10000000
## 3685 3685                            Eminent Technology Solutions   0.80   6300000
```

```
## 3704 3704              Infinity Consulting Solutions      0.79  35700000
## 3716 3716                         LISS Consulting         0.79   2400000
## 3764 3764      American Portfolios Financial Services     0.77 117000000
## 3787 3787                        Optimax Systems          0.76  20100000
## 3789 3789                           Xtreme Trips          0.76   6700000
## 3864 3864                         ClassBook.com           0.73   9600000
## 3871 3871                     eTransMedia Technology       0.72  25900000
## 3891 3891                             Net@Work            0.72  29300000
## 3899 3899                    Denihan Hospitality Group     0.71 280800000
## 3924 3924             Finger Lakes Technologies Group      0.70  23100000
## 3925 3925                                  Sinu           0.70   3000000
## 3943 3943                          Gersh Academy          0.70   8600000
## 4000 4000                          Send Word Now          0.67  16600000
## 4003 4003                      Ovation Travel Group        0.67  58600000
## 4007 4007                     Spear Physical Therapy       0.67   5200000
## 4008 4008                          Nutricap Labs          0.67  30000000
## 4039 4039                          Empire Office          0.66 349600000
## 4054 4054                                  BUMI           0.65   2300000
## 4117 4117                            NorthPoint           0.64  27300000
## 4128 4128                     Cvision Technologies         0.63   5200000
## 4153 4153                        Everyday Health          0.62 146400000
## 4154 4154         Business Communications Management       0.62  10900000
## 4158 4158                             Edulence            0.62   2000000
## 4162 4162                           JR Products          0.62   4900000
## 4166 4166                        Motivators.com          0.62  13200000
## 4170 4170             Creative Environment Solutions       0.62   5300000
## 4224 4224                             Fragomen            0.60 346500000
## 4245 4245             Computer Resources of America        0.59   7700000
## 4250 4250                             NiCoForm            0.59   2100000
## 4287 4287                            LinguaLinx           0.57   4300000
## 4307 4307                            ItsHot.com           0.57   5700000
## 4335 4335                       New York Kids Club         0.56  12500000
## 4336 4336                        Cyrus Innovation         0.56   5900000
## 4352 4352                              Kitware           0.55  21300000
## 4363 4363                           TransPerfect          0.55 341300000
## 4399 4399                          Direct Agents          0.54  22700000
## 4422 4422                        BPA International         0.53   5400000
## 4432 4432                          C2C Outdoor           0.53  27700000
## 4447 4447                     Select Office Suites         0.52   7600000
## 4448 4448                       The Atlantic Group         0.52  64300000
## 4465 4465                             OpenLink            0.52 304300000
## 4467 4467                         Blue Telescope         0.52   3400000
## 4471 4471                    KJ Technology Consulting      0.51   4000000
## 4474 4474                    Sam Schwartz Engineering      0.51  16800000
## 4475 4475                    Presentation Products        0.51   9700000
## 4490 4490                       Catseye Pest Control       0.51   9000000
## 4522 4522                        Geneva Worldwide         0.50   7400000
## 4535 4535                     Capital Access Network       0.49 151600000
## 4544 4544              ePromos Promotional Products        0.49  28400000
## 4552 4552                             Paradysz            0.49  41000000
## 4559 4559                              Tabush           0.48   3900000
## 4560 4560          Sentry Communications & Security        0.48   6000000
## 4565 4565                           EquipSystems          0.48   3000000
## 4569 4569                        Bogota Latin Bistro       0.48   3800000
```

```
## 4595 4595                                                    Numerix   0.48  56400000
## 4602 4602                                            Brite Computers   0.47  19900000
## 4605 4605                                               QED National   0.47   6700000
## 4611 4611                                                    Inergex   0.47  17400000
## 4632 4632                                                Transfinder   0.47   7800000
## 4639 4639  American Petroleum Equipment & Construction              0.46  10400000
## 4645 4645                                               DatapointLabs   0.46   2500000
## 4646 4646                                                  Mimeo.com   0.46  87700000
## 4652 4652                                                 SaveAround   0.46   6900000
## 4670 4670                                                    VarData   0.45  14800000
## 4671 4671                                                  Webucator   0.45   4600000
## 4702 4702                     Copier Fax Business Technologies        0.44   5300000
## 4727 4727                                             Arrow Security   0.44  14000000
## 4728 4728                                         DurAmerica Brokerage   0.43   2700000
## 4731 4731                                                 PlayScripts   0.43   2600000
## 4732 4732                                                    Mycroft   0.43  26400000
## 4738 4738                                                 McMurry/TMG   0.43  91400000
## 4747 4747                                                  TravelClick   0.43 257100000
## 4755 4755                                             Greyston Bakery   0.42  10100000
## 4774 4774                                             The Lloyd Group   0.42   9900000
## 4782 4782                     Life Safety Engineered Systems          0.42   7100000
## 4800 4800                                     Documentation Strategies   0.41   2300000
## 4803 4803                                            Benefit Resource   0.41  13300000
## 4820 4820                                                Aluf Plastics   0.40 102900000
## 4839 4839                                                   Makovsky   0.40  14300000
## 4849 4849                                        GTM Payroll Services   0.40   6700000
## 4853 4853                                                     Affect   0.39   2600000
## 4891 4891                              Express Air Freight Unlmited    0.38  28600000
## 4896 4896                     Rand Engineering & Architecture          0.38   9800000
## 4913 4913                                              Jackson Lewis   0.37 352000000
## 4934 4934                                             Buffalo Filter   0.37  15000000
## 4944 4944                                              McElroy Deutsch   0.36 116500000
## 4950 4950                     Colorspec Coatings International          0.36   3300000
## 4954 4954                                         Premium Productions   0.36   6800000
## 4981 4981  SmartSource Computer & Audio Visual Rentals               0.35  57300000
##                                      Industry Employees          City State
## 26      Consumer Products & Services          17       New York    NY
## 30           Advertising & Marketing          79       New York    NY
## 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
## 71           Advertising & Marketing          42       New York    NY
## 124              Telecommunications          28       New York    NY
## 126                          Retail          17        Hewlett    NY
## 153           Travel & Hospitality          42       Brooklyn    NY
## 174                           Media          99       New York    NY
## 218                     IT Services         119       Fishkill    NY
## 231                       Education          22       New York    NY
## 249          Advertising & Marketing          15       New York    NY
## 263  Business Products & Services           4       New York    NY
## 266                           Media           4       New York    NY
## 274              Telecommunications         213         Walden    NY
## 279                           Media           7       New York    NY
```

```
## 284      Advertising & Marketing    46        New York      NY
## 298             Manufacturing       12      Ronkonkoma      NY
## 308           Human Resources       30        New York      NY
## 313  Business Products & Services   48        New York      NY
## 387         Telecommunications       6          colden      NY
## 399                    Health        6        New York      NY
## 417      Advertising & Marketing    19        New York      NY
## 449      Advertising & Marketing    45        Brooklyn      NY
## 464  Consumer Products & Services   25        Brooklyn      NY
## 489         Telecommunications      25        New York      NY
## 496          Financial Services    198        New York      NY
## 501      Advertising & Marketing    12        New York      NY
## 514         Telecommunications      31          Latham      NY
## 520  Consumer Products & Services   51        New York      NY
## 521            Food & Beverage      56  Pleasant Valley     NY
## 525                     Media      14        New York      NY
## 567      Advertising & Marketing    25        new york      NY
## 608                    Energy     137   Ballston Lake      NY
## 631                    Health       2        New York      NY
## 655          Financial Services     70        New York      NY
## 659          Financial Services    153        New York      NY
## 660                    Health     121     Floral Park      NY
## 668                     Media      46        New York      NY
## 700                     Media     200        New York      NY
## 708      Advertising & Marketing    28        Melville      NY
## 713  Consumer Products & Services   18        Brooklyn      NY
## 716          Financial Services    102 Long Island City     NY
## 727      Advertising & Marketing    12        New York      NY
## 743                     Retail     22        Brooklyn      NY
## 749                     Media     146        New York      NY
## 751                 Education     100        New York      NY
## 768            Food & Beverage      14        Brooklyn      NY
## 787  Consumer Products & Services   16        Brooklyn      NY
## 808               Real Estate      15          Goshen      NY
## 814      Advertising & Marketing    51 Saratoga Springs     NY
## 817                IT Services      15        Scarsdale      NY
## 829      Advertising & Marketing    40        New York      NY
## 831       Travel & Hospitality       6        New York      NY
## 853               Real Estate      30        New York      NY
## 882                IT Services     142          Ithaca      NY
## 889          Financial Services    207        New YOrk      NY
## 895      Advertising & Marketing     8        new york      NY
## 921                     Retail     75        New York      NY
## 934                    Energy      90         Oneonta      NY
## 970                 Education      19   New York City      NY
## 1009     Advertising & Marketing    28        New York      NY
## 1019          Human Resources       12        Fairport      NY
## 1027 Logistics & Transportation      1      Farmingdale      NY
## 1035               IT Services      40        New York      NY
## 1047                    Media      45        New York      NY
## 1053 Consumer Products & Services  100        new york      NY
## 1066     Advertising & Marketing    22      West Islip      NY
## 1068               IT Services     382         Syracuse      NY
## 1090     Advertising & Marketing   180        New York      NY
```

```
## 1094         Advertising & Marketing       2         New York    NY
## 1099 Consumer Products & Services          20       Garden City    NY
## 1118              Manufacturing            40         New York    NY
## 1128         Advertising & Marketing      65         New York    NY
## 1165 Business Products & Services          30         New York    NY
## 1180         Advertising & Marketing      28         New York    NY
## 1189               IT Services           295         New York    NY
## 1202               IT Services            27         New York    NY
## 1209         Advertising & Marketing      15        Great Neck    NY
## 1213              Manufacturing           11         New York    NY
## 1233                    Retail            11          ASTORIA     NY
## 1258                    Retail            61         New York    NY
## 1292         Advertising & Marketing      51         New York    NY
## 1370                  Software          220         New York    NY
## 1381                 Education            40         New York    NY
## 1422         Advertising & Marketing      87         New York    NY
## 1441                 Insurance            50          St James    NY
## 1442           Human Resources             7         New York    NY
## 1450                    Health            37      White Plains    NY
## 1457                    Retail            12          Brooklyn    NY
## 1484        Government Services           17            Nyack     NY
## 1498           Human Resources          2081         New York    NY
## 1502         Advertising & Marketing      38        IRVINGTON     NY
## 1539                    Retail            15           Inwood     NY
## 1558                  Software           25         New York    NY
## 1564         Advertising & Marketing     250         New York    NY
## 1608         Advertising & Marketing      35         New York    NY
## 1611           Human Resources             8         New York    NY
## 1621  Logistics & Transportation          22          Melville    NY
## 1639               IT Services           500         New York    NY
## 1645               IT Services            23         Rochester    NY
## 1705         Advertising & Marketing      41         New York    NY
## 1708 Consumer Products & Services           5       Lindenhurst    NY
## 1716 Consumer Products & Services          10 Long Island City    NY
## 1722              Construction            20           Buffalo    NY
## 1754               IT Services            47           Latham     NY
## 1766 Business Products & Services          15           Monsey     NY
## 1875            Food & Beverage           41         Canastota    NY
## 1876         Telecommunications           22          Fairport    NY
## 1878              Construction            25            Bronx     NY
## 1894               IT Services           237      Clifton Park    NY
## 1919         Financial Services           14   Oakland Gardens    NY
## 1953                 Education            25         New York    NY
## 1981                 Education           100         New York    NY
## 1984                   Energy            120         Getzville    NY
## 1985                    Retail             3          Brooklyn    NY
## 2066              Manufacturing           77         New York    NY
## 2076 Consumer Products & Services          23         New York    NY
## 2088         Advertising & Marketing      15         New York    NY
## 2115         Financial Services           33         New York    NY
## 2141       Travel & Hospitality           61        Ronkonkoma    NY
## 2172                    Health           181      Lake Success    NY
## 2178                    Retail            75       Cheektowaga    NY
## 2197 Business Products & Services          32     Honeoye Falls    NY
```

```
## 2208    Advertising & Marketing    44         New York  NY
## 2217                    Software   320         New York  NY
## 2233                IT Services    13          Holbrook  NY
## 2279                    Software    80     Staten Island  NY
## 2291                    Security    25        Massapequa  NY
## 2335             Human Resources    56          New York  NY
## 2351 Business Products & Services   12          New York  NY
## 2371    Advertising & Marketing    32          New York  NY
## 2398    Advertising & Marketing    74          New York  NY
## 2404    Advertising & Marketing    63          New York  NY
## 2431               Manufacturing    13           Buffalo  NY
## 2435                IT Services   300          New York  NY
## 2451            Food & Beverage    30          Brooklyn  NY
## 2474                IT Services    52         Rochester  NY
## 2493 Business Products & Services  106          New York  NY
## 2502                      Retail    12          New York  NY
## 2505                IT Services    46         Scarsdale  NY
## 2525          Telecommunications    85          Fairport  NY
## 2541                      Health    50             Utica  NY
## 2548                      Health    20          New York  NY
## 2555                      Energy   294       New Windsor  NY
## 2560                 Real Estate     7           Syosset  NY
## 2561          Telecommunications    15    Roslyn Heights  NY
## 2584                IT Services    14          New York  NY
## 2621 Business Products & Services   11          Brooklyn  NY
## 2655                IT Services    26          New York  NY
## 2665            Food & Beverage    12           Menands  NY
## 2674                      Energy     5          New York  NY
## 2677                      Health   166             troy  NY
## 2701               Manufacturing    26          blauvelt  NY
## 2705    Advertising & Marketing    15     Pleasantville  NY
## 2717    Advertising & Marketing    32          New York  NY
## 2746          Telecommunications    20          New York  NY
## 2750    Advertising & Marketing   133          New York  NY
## 2763 Consumer Products & Services    9          New York  NY
## 2783                    Software    21 Saratoga Springs  NY
## 2794                      Health    67          New York  NY
## 2798                   Education    21     White Plains  NY
## 2804    Advertising & Marketing   131          New York  NY
## 2827                IT Services   125         Rochester  NY
## 2829 Business Products & Services 1124          New York  NY
## 2830    Advertising & Marketing    35          New York  NY
## 2876           Computer Hardware    44          New York  NY
## 2888   Logistics & Transportation   25           Amherst  NY
## 2895          Telecommunications   316          New York  NY
## 2923 Business Products & Services   90     White Plains  NY
## 2963                    Software   800          New York  NY
## 2970          Financial Services   483          Fairport  NY
## 2980                IT Services    95         Hauppauge  NY
## 2994             Human Resources  1000          New York  NY
## 2996          Telecommunications   250         ROCHESTER  NY
## 3008          Telecommunications   268         Rochester  NY
## 3020                IT Services   322          New York  NY
## 3053                Construction   219         Rochester  NY
```

```
## 3055      Advertising & Marketing     10        New York    NY
## 3059                       Retail      10        New York    NY
## 3096 Consumer Products & Services      64        Henrietta   NY
## 3102 Business Products & Services      34        Brooklyn    NY
## 3114                       Retail      15        New York    NY
## 3116      Advertising & Marketing      37        New York    NY
## 3135             Human Resources     1134        New York    NY
## 3145      Advertising & Marketing      21        New York    NY
## 3152      Advertising & Marketing     186        New York    NY
## 3167             Human Resources     107        New York    NY
## 3192               Manufacturing      22        Freeport    NY
## 3209                    Education      50        New York    NY
## 3250      Advertising & Marketing       9        New York    NY
## 3261                  IT Services     210        Melville    NY
## 3288      Advertising & Marketing      73        Rochester   NY
## 3312                        Media      16        New York    NY
## 3350                    Education      70        New York    NY
## 3354 Consumer Products & Services      55        Brooklyn    NY
## 3357             Human Resources      53        Fairport    NY
## 3382                    Education      55     Port Chester   NY
## 3386                  IT Services    1000        New York    NY
## 3400                  IT Services      80        New York    NY
## 3423                  IT Services      39 Long Island City   NY
## 3500             Food & Beverage     383        New York    NY
## 3532                 Construction      24        Woodhaven   NY
## 3535 Consumer Products & Services      87          Albany    NY
## 3542             Food & Beverage       5        New York    NY
## 3544                  IT Services      24        New York    NY
## 3549      Advertising & Marketing      24        New York    NY
## 3584                  IT Services     432        New York    NY
## 3592               Manufacturing     132          Bergen    NY
## 3594      Advertising & Marketing      32        New York    NY
## 3601 Consumer Products & Services      85      Schenectady   NY
## 3603           Financial Services      35        New York    NY
## 3632                  IT Services     144        Tarrytown   NY
## 3643 Business Products & Services      45        New York    NY
## 3661       Environmental Services     250        Syracuse    NY
## 3663                  IT Services      33        Brooklyn    NY
## 3685                  IT Services      48          Buffalo   NY
## 3704             Human Resources     325        New York    NY
## 3716                  IT Services      18     Lake Success   NY
## 3764           Financial Services      81        Holbrook    NY
## 3787               Manufacturing     200          Ontario   NY
## 3789        Travel & Hospitality      30        plainview   NY
## 3864                    Education      61        Castleton   NY
## 3871                       Health     298            Troy    NY
## 3891                  IT Services     156        New York    NY
## 3899        Travel & Hospitality    2280        New York    NY
## 3924           Telecommunications     105          Victor    NY
## 3925                  IT Services      25        New York    NY
## 3943                    Education      51       Huntington   NY
## 4000           Telecommunications      59        New York    NY
## 4003        Travel & Hospitality     509        New York    NY
## 4007                       Health      40        New York    NY
```

```
## 4008                       Health    45   Farmingdale   NY
## 4039 Business Products & Services   345     New York   NY
## 4054                  IT Services     8     New York   NY
## 4117 Business Products & Services   144     New York   NY
## 4128                     Software    28   Forest Hills  NY
## 4153                        Media   602     New York   NY
## 4154           Telecommunications    21     New York   NY
## 4158                     Software    15     New York   NY
## 4162                Manufacturing    30  Clarence Center NY
## 4166        Advertising & Marketing  55      westbury   NY
## 4170        Environmental Services   60     New York   NY
## 4224 Business Products & Services   404     New York   NY
## 4245                  IT Services    47     New York   NY
## 4250                Manufacturing    16     Rochester   NY
## 4287 Business Products & Services    32       Cohoes   NY
## 4307                       Retail     9     New York   NY
## 4335                    Education   200     New York   NY
## 4336                     Software    46     new york   NY
## 4352                     Software   110   Clifton Park  NY
## 4363 Business Products & Services  2218     New York   NY
## 4399       Advertising & Marketing   47     New York   NY
## 4422           Telecommunications   139   Carle Place   NY
## 4432       Advertising & Marketing   12     New York   NY
## 4447                  Real Estate    21     New York   NY
## 4448 Business Products & Services    51     New York   NY
## 4465                     Software  1271     Uniondale   NY
## 4467                        Media     9     New York   NY
## 4471                  IT Services    13     New York   NY
## 4474                  Engineering    94     New York   NY
## 4475                  Engineering    29     New York   NY
## 4490 Consumer Products & Services    62     Castleton   NY
## 4522 Business Products & Services    30     new york   NY
## 4535           Financial Services   425     New York   NY
## 4544       Advertising & Marketing  104     New York   NY
## 4552       Advertising & Marketing  270     New York   NY
## 4559                  IT Services    23     New York   NY
## 4560                     Security    40    Hicksville   NY
## 4565                       Health    31     New York   NY
## 4569              Food & Beverage    63      Brooklyn   NY
## 4595                     Software   198     New York   NY
## 4602                  IT Services    54        Victor   NY
## 4605                  IT Services    89     New York   NY
## 4611                  IT Services   140       Buffalo   NY
## 4632                     Software    63   Schenectady   NY
## 4639                 Construction    68        Walden   NY
## 4645                  Engineering    11        Ithaca   NY
## 4646 Business Products & Services   632     New York   NY
## 4652       Advertising & Marketing   64    Binghamton   NY
## 4670           Telecommunications    18     Rochester   NY
## 4671                    Education    24    Jamesville   NY
## 4702 Business Products & Services    27       Buffalo   NY
## 4727                     Security   450     New York   NY
## 4728                    Insurance    15 Long Island City NY
## 4731                       Retail    10     New York   NY
```

```
## 4732                    IT Services      275      New York    NY
## 4738      Advertising & Marketing        269   New York City   NY
## 4747      Travel & Hospitality           906      New York    NY
## 4755           Food & Beverage            84        Yonkers    NY
## 4774                    IT Services       73      New York    NY
## 4782                       Security       25        buffalo    NY
## 4800                    IT Services       25     Rensselaer    NY
## 4803 Business Products & Services         96      Rochester    NY
## 4820                  Manufacturing      307     Orangeburg    NY
## 4839      Advertising & Marketing         55      New York    NY
## 4849           Financial Services         43    Clifton Park   NY
## 4853      Advertising & Marketing         18      New York    NY
## 4891   Logistics & Transportation        70        Jamaica    NY
## 4896                    Engineering       80      New York    NY
## 4913 Business Products & Services        690   White Plains    NY
## 4934                  Manufacturing       67      Lancaster    NY
## 4944 Business Products & Services        288      New York    NY
## 4950                   Construction       10       Holbrook    NY
## 4954      Advertising & Marketing         16   Pleasantville   NY
## 4981 Business Products & Services        296      Hauppauge    NY
```

```r
# dimensions after
dim_outlier <-as.numeric(dim(inc_ny)[1])
dim_no_outlier <- as.numeric(dim(inc_ny_no_outlier)[1])

# compare records of dataset
print(paste0("Rows prior to outlier reduction ", dim_outlier))
```

```
## [1] "Rows prior to outlier reduction 311"
```

```r
print(paste0("Rows after 1% outlier reduction ",dim_no_outlier))
```

```
## [1] "Rows after 1% outlier reduction 308"
```

Defining outliers with the classic

$$Lower = Q1 - (1.5 * IQR) \quad Upper = Q1 + (1.5 * IQR)$$

removes 45 rows which is too drastic, so I chose to remove only 1% from top and bottom which only removes the top 3 rows. Different quantiles could be chosen as well.

```r
# groupby and plot without outliers
groupby_column(inc_ny_no_outlier, Industry, Employees) %>%
  crossbar_plot(Industry, med_grp, med_grp, min_grp, max_grp,
            title = "Median Employee in NY State by Industry (No Outliers)")
```

Median Employee in NY State by Industry (No Outliers)
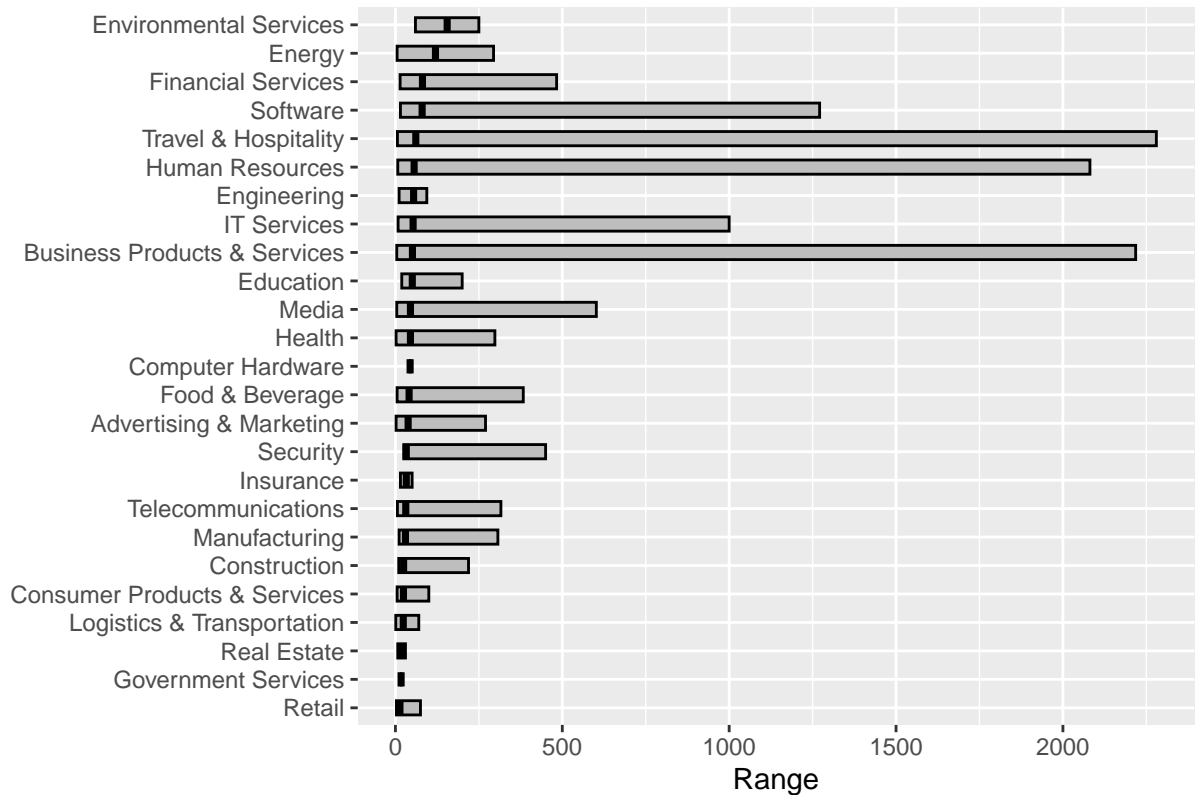
This showcases the median and range of `Employees` by `Industry` in NY state. That said, it is important to understand which records were removed. If I had more time I would compute that and experiment with more quantiles.

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.

```
# group by Industry then get revenue per emp
(inc_ny_rev_emp <- inc_ny %>%
  group_by(Industry) %>%
  summarise(n = n(),
            num_emp = sum(Employees),
            rev_sum = sum(Revenue)) %>%
  mutate(rev_emp_per_1k = round((rev_sum / num_emp) / 1000)) %>%
  select(Industry, n, rev_emp_per_1k) %>%
  arrange(desc(rev_emp_per_1k)))
```

```
## # A tibble: 25 x 3
##    Industry                     n rev_emp_per_1k
##    <chr>                    <int>          <dbl>
##  1 Energy                       5            650
##  2 Logistics & Transportation   4            637
```

```
##  3 IT Services                     43        550
##  4 Computer Hardware                1        520
##  5 Insurance                        2        474
##  6 Retail                          14        473
##  7 Consumer Products & Services    17        451
##  8 Financial Services              13        404
##  9 Telecommunications              17        387
## 10 Manufacturing                   13        386
## # ... with 15 more rows
```

```
inc_ny_rev_emp %>%
  ggplot() +
    geom_bar(aes(rev_emp_per_1k, reorder(Industry, rev_emp_per_1k)), stat = "Identity")
```