Data Science Project

KBDio

5/30/2021

Abstract

This project deals with various visualization techniques applied to a dataset to make it easy to draw conclusions about the data by looking at the plots. This project aims to explore information about Service Express (SEI) service, their customers and equipment they maintain. The SEI data consist of six different datasets that are linked together by various variables such as unique keys and item numbers. Finding the right questions and significant variables are important to unveil hidden information in this project. The results show that SEI have strong customer relations in the Midwest region. The challenging part of this project is the cleaning and exploring insights with little to no additional information about the company.

Introduction

This project is focus on using statistical graphics and data visualization methods to convey information about service Express services.The three main objective of this work is to

- Explore the Customer base
- Explore Service Express Service
- Explore information about their equipment maintenance

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
## here() starts at C:/Users/Kanyanee/Desktop/compettition
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
## Loading required package: sp
```

```
## Checking rgeos availability: TRUE
```

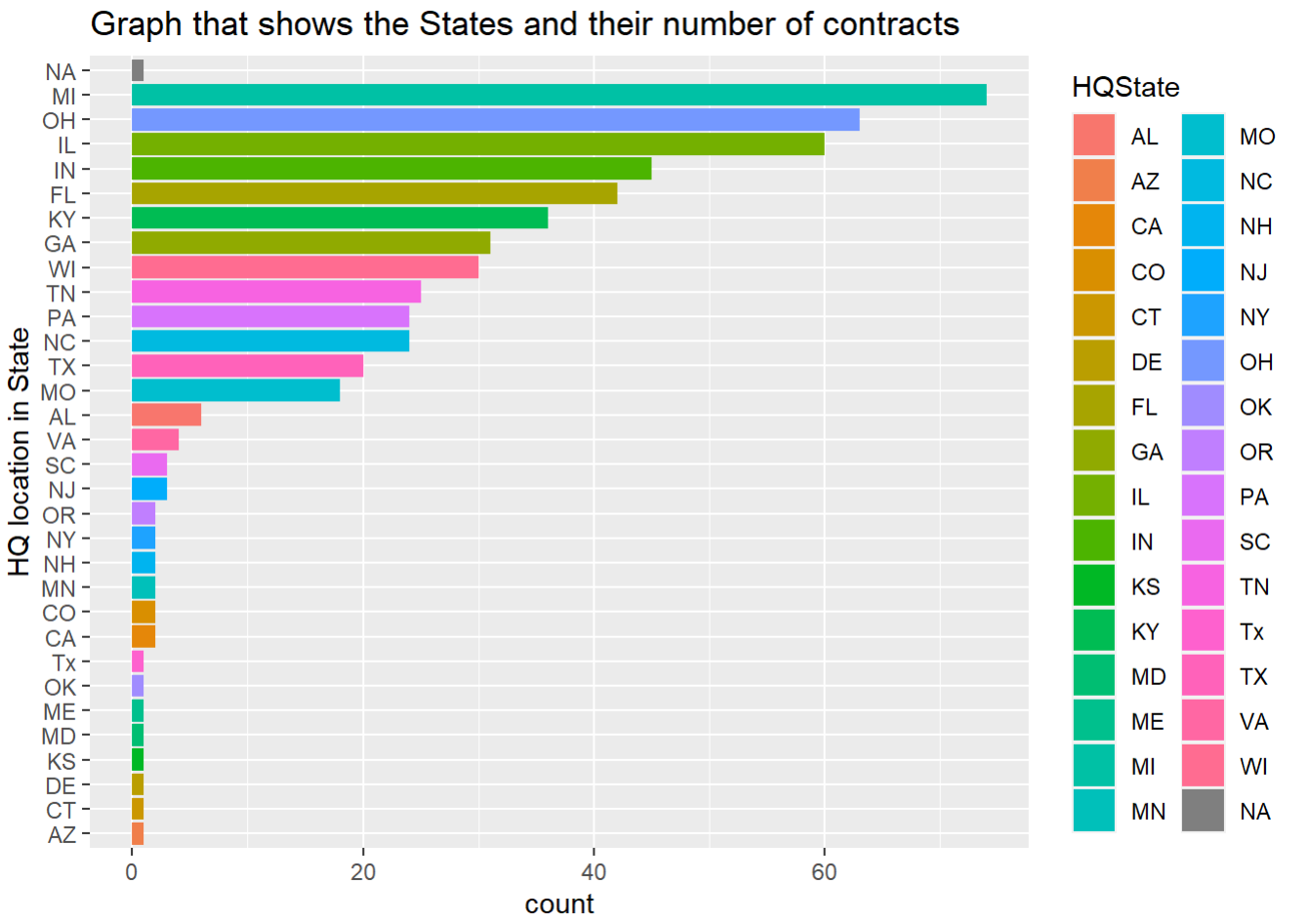
```
## GDAL version >= 3.1.0 | setting mapviewOptions(fgb = TRUE)
```

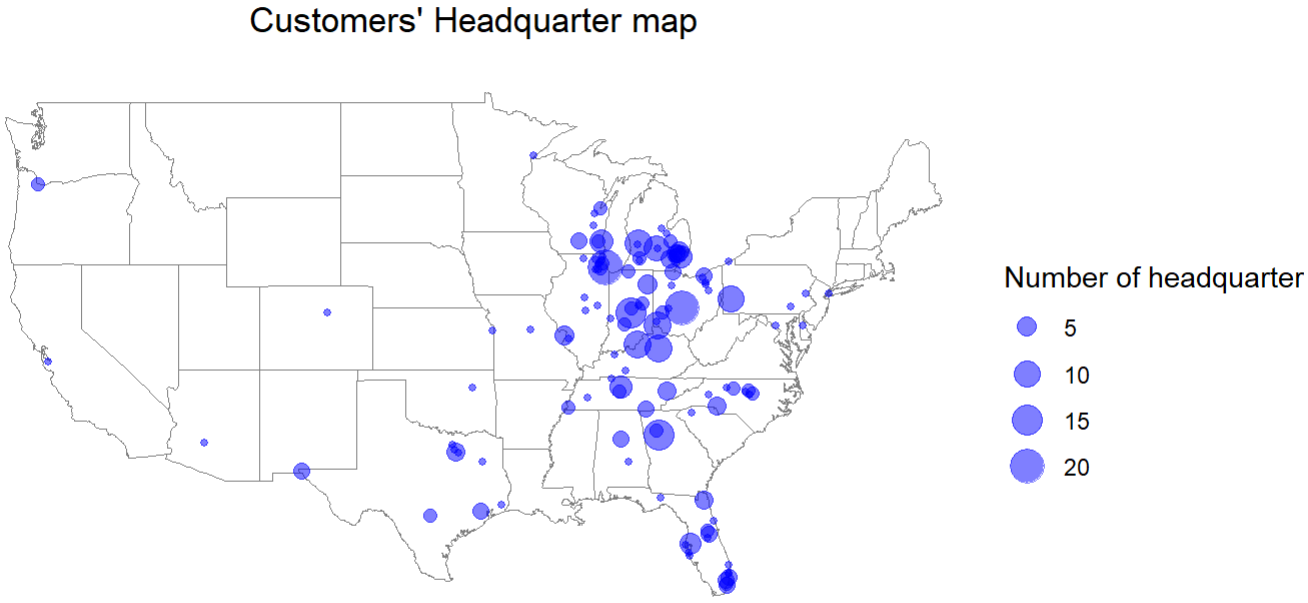
Loading and cleaning of data

For good visualization of data, the data must be cleaned before we plot any of the graphs, otherwise the plots would be displaying inaccurate/false information. The plot must show accurate information as they are used to draw important conclusion about the data. There are three areas of the data that needs tidying up and these are removing of all extra columns with missing values, removing the missing values at the latter rows and renaming, respelling and recoding of column variables and values.

Implementation of graphs to explore more on Customer Base

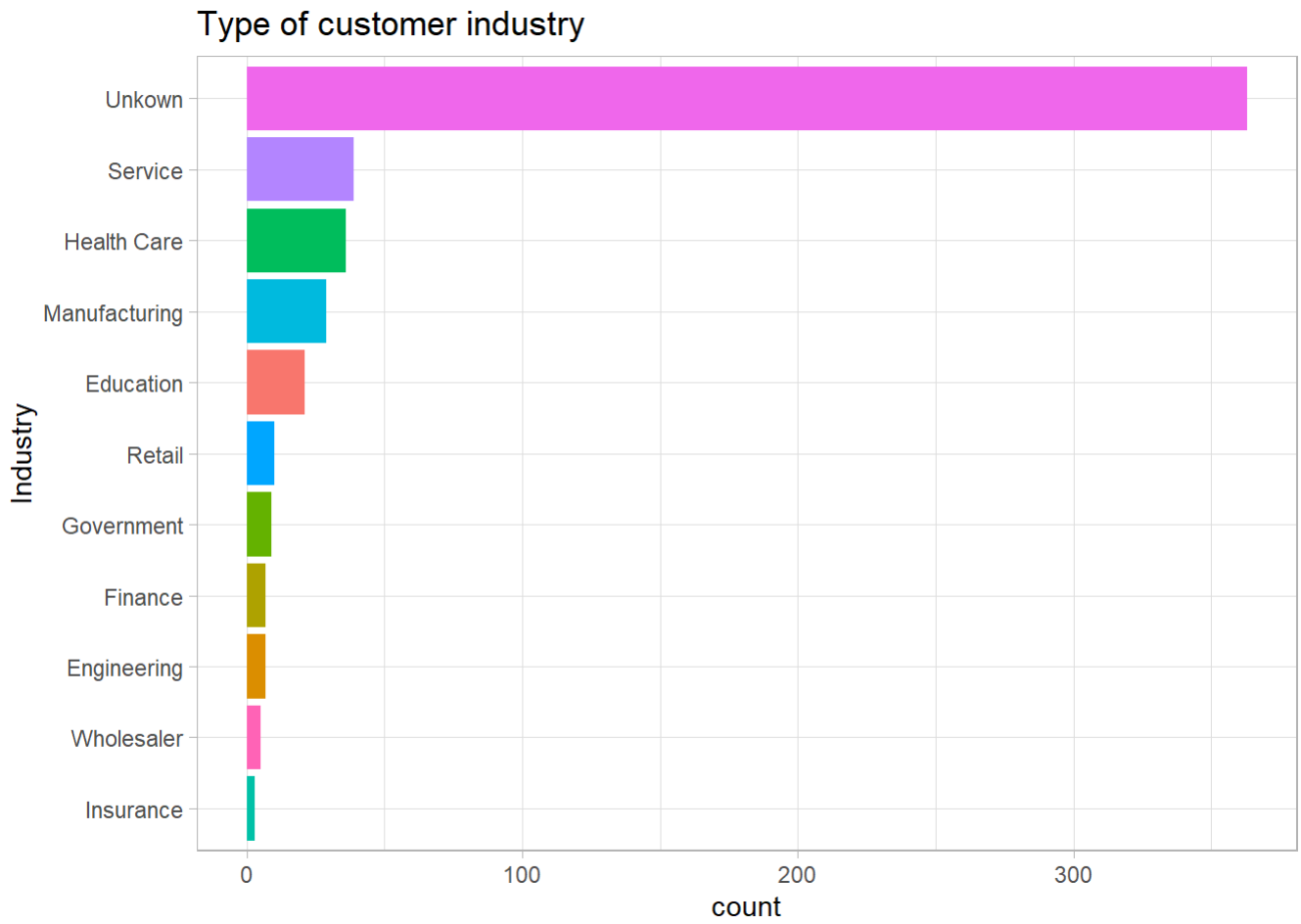
Headquarters location:From the data, we grouped the various headwaters of the company according to their location to find the location or state with the most number of SEI express service offices.From the exploratory analysis, we found out that Michigan had the most number of Head QuaRters offices of about 74 branches followed by OH, IL, IN. This outcome is displayed in the bar graph and the map below.


Graph that shows the States and their number of contracts
HQ location in State
NA
MI
OH
IL
IN
FL
KY
GA
WI
TN
PA
NC
TX
MO
AL
VA
SC
NJ
OR
NY
NH
MN
CO
CA
Tx
OK
ME
MD
KS
DE
CT
AZ
0
20
40
60
count
HQState
AL
AZ
CA
CO
CT
DE
FL
GA
IL
IN
KS
KY
MD
ME
MI
MN
MO
NC
NH
NJ
NY
OH
OK
OR
PA
SC
TN
Tx
TX
VA
WI
NA

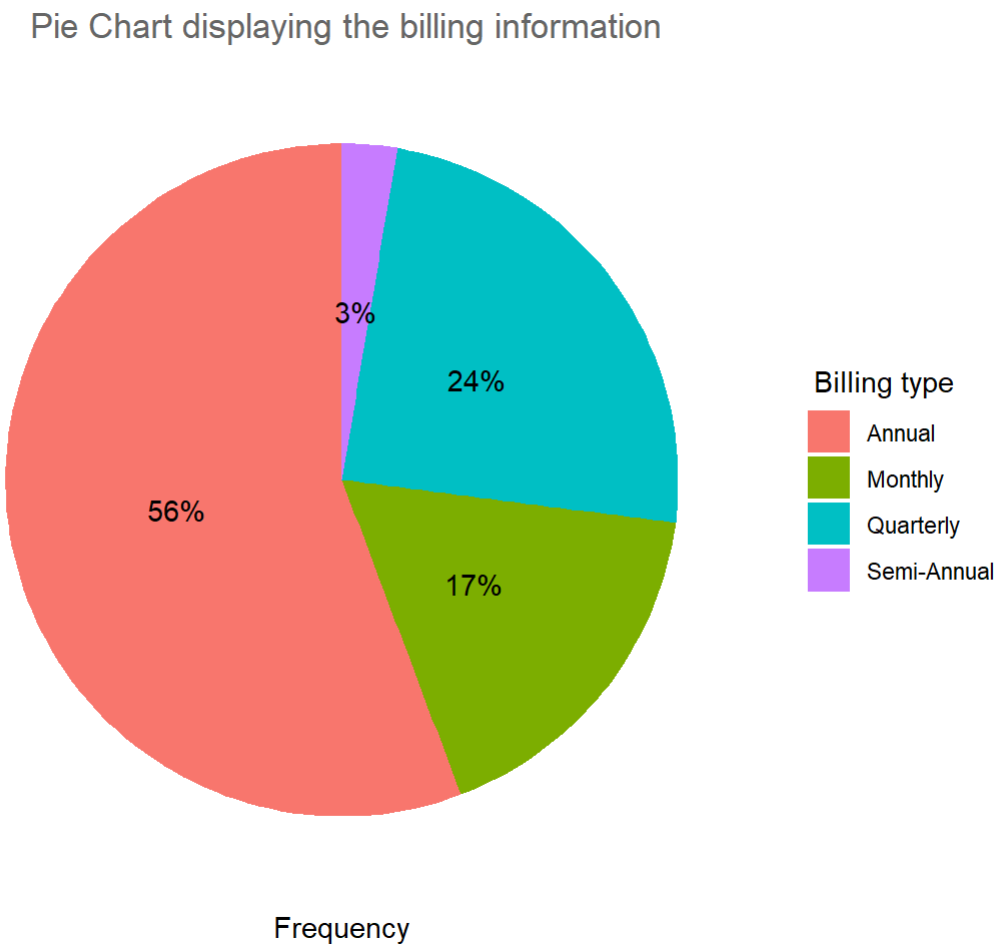

Customers' Headquarter map
Number of headquarter
5
10
15
20

Industry: One of the primary goals of most companies is to provide quality services and expand their operations across the states. In this work, we perform exploratory analysis to identify the various industries we provide services according to their number of operation. From the table, we realized that the unknown category of the industry appear to have the most frequency followed by Service, Health care, Manufacturing, Education, Retail, Government, Engineering, Finance, Wholesale and insurance.

Industry	count
Unkown	363
Service	39
Health Care	36
Manufacturing	29
Education	21
Retail	10
Government	9
Engineering	7
Finance	7
Wholesaler	5
Insurance	3



Billing analysis:The plot below shows the different billing ways customers use to pay for their services rendered by SEI Express service. The bar plot displays the categorize of the billing types according to their frequency.The plot shows the frequency on the y-axis and the billing type on the x-axis. From the plot, we can see that the annual billing is the most preferred option, followed by quarterly, monthly and semi-annual in order of decreasing.

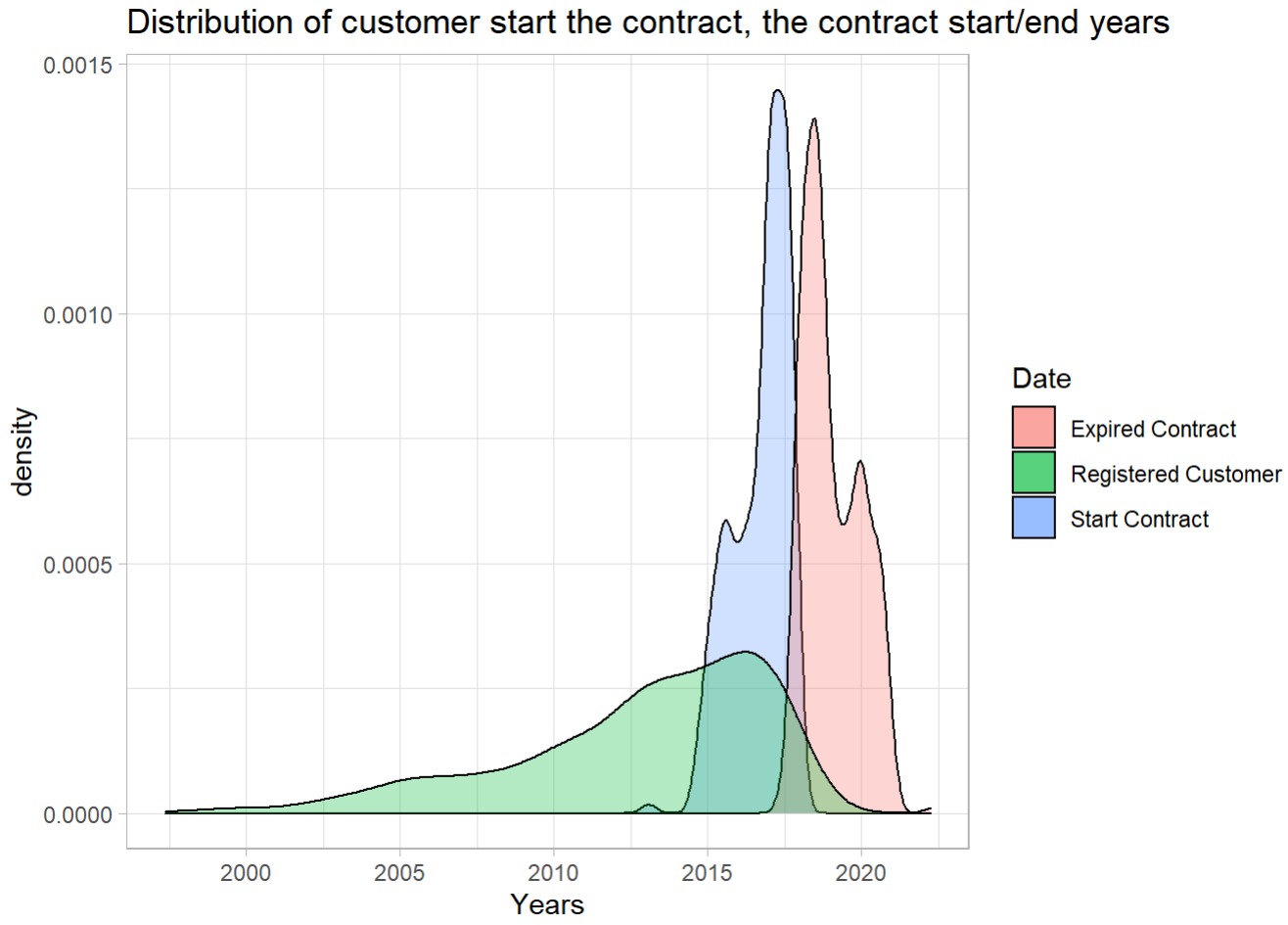


Contract Duration: The duration of contract secure SEI with income and job offer to continue the business. From the table below, we compared various contracts using their start and expired dates to understand difference that exist in their contract duration. The goal is to know the customer whose contract existed the longest.One of the objective of every company is to make profit and with the kind of services we render to our customer, the more contract length we have with our customers the more profit we make. Therefore, it is important to pay more attention to customers with longer contracts. The duration will help SEI to plan its business strategy for the company growth. Customer's AgrPKey number 3623, 2078, 10905, and 10117 make the longest contract with SEI at approximately 5 years.

AgrPkey	ContractStartDate	ContractExpirationDate	Difference	Durration
3623	2014-11-01	2019-10-31	1825 days	157680000s (~5 years)
2078	2013-01-01	2017-12-31	1825 days	157680000s (~5 years)
10905	2013-03-01	2018-02-28	1825 days	157680000s (~5 years)
10117	2017-04-01	2022-03-31	1825 days	157680000s (~5 years)
7986	2016-01-01	2019-12-31	1460 days	126144000s (~4 years)
13821	2016-01-01	2019-12-31	1460 days	126144000s (~4 years)
17525	2017-01-01	2020-06-30	1276 days	110246400s (~3.49 years)
11307	2014-11-21	2017-12-31	1136 days	98150400s (~3.11 years)
13415	2014-12-16	2017-12-31	1111 days	95990400s (~3.04 years)

AgrPkey	ContractStartDate	ContractExpirationDate	Difference	Durration
2087	2016-01-01	2018-12-31	1095 days	94608000s (~3 years)

Distribution of customer and their contract duration : The plot shows the distribution of the date of customer registration, start of contract and contract expiration. As seen from the density plot below, there are to peaks of date start contract and expired contract of customers of SEI express Service. The first peak, represent the start date of contracts and it is evident that most contracts were issued or started around 2017.The second peak also represent the expiring date of customer's contracts, and most of the contracts ended in 2018.



SEI service

Map of ticket calls by state

Knowing the service location is important for operation and enhancing the service. This will allow SEI to reach customer and provide its service in time without pausing production or service. A Heat map is used in this analysis to easily visualize the location and the volume of ticket events. The heat map of number of service ticket by state show that The volume of ticket event is significantly high in Indiana, Illinois, Ohio, and Michigan.

3. plot the map

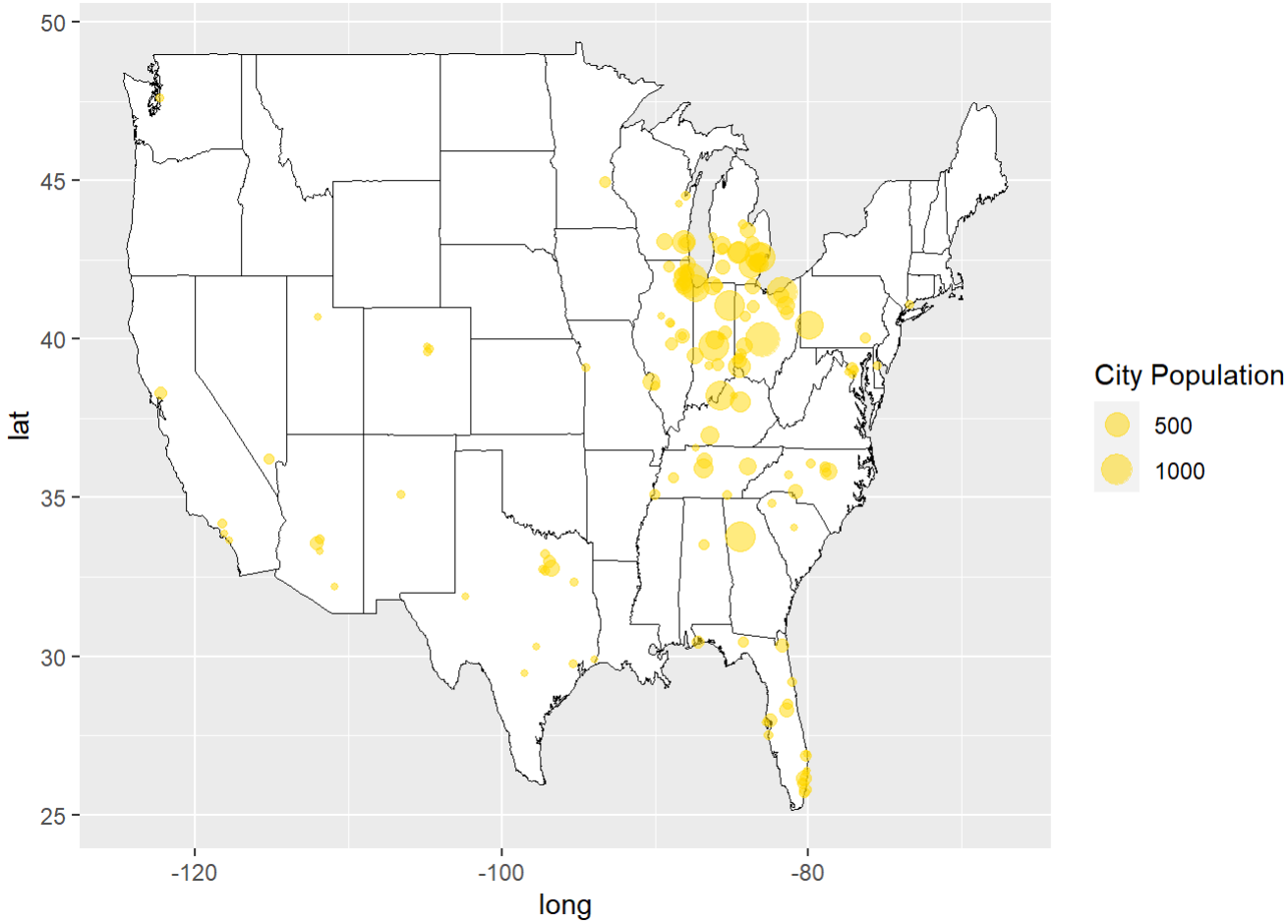
#get the map and convert it to a SpatialPolygons: mapUSA <- map('state', fill = TRUE, plot = FALSE) nms <- sapply(strsplit(mapUSA$names, ':'), function(x)x[1]) USApolygons <- map2SpatialPolygons(mapUSA, IDs = nms, CRS('+proj=longlat'))

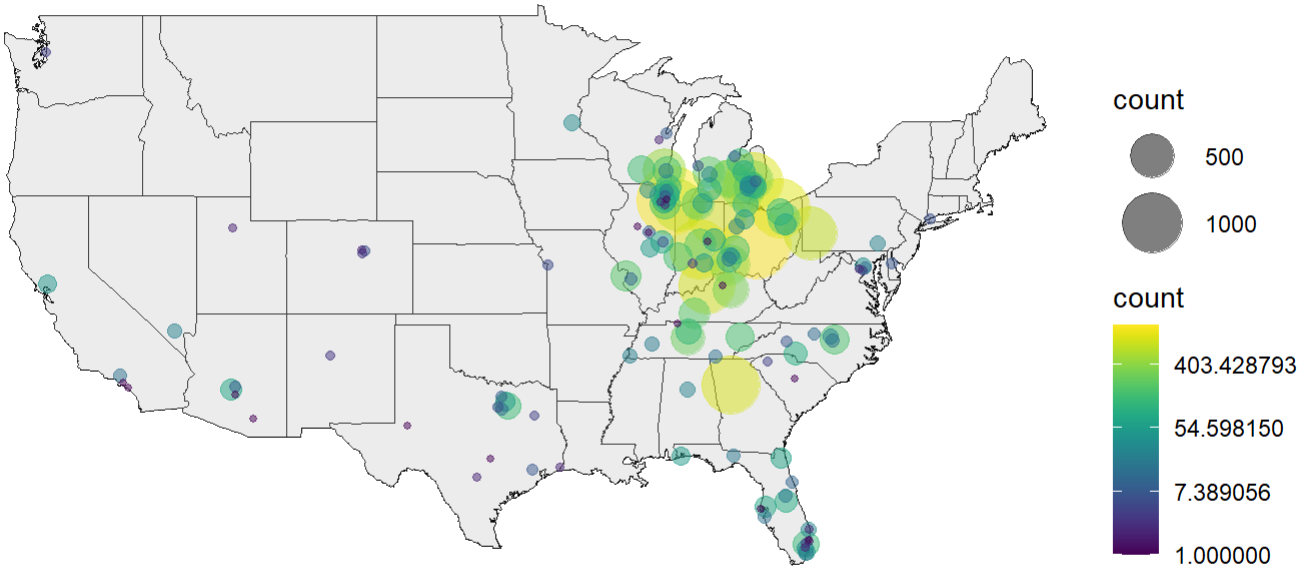
#now add the information from your data: idx <- match(unique(nms), dat
$states)dat2 < -data.frame(count = dat$count[idx], state = unique(nms)) row.names(dat2) <- unique(nms)

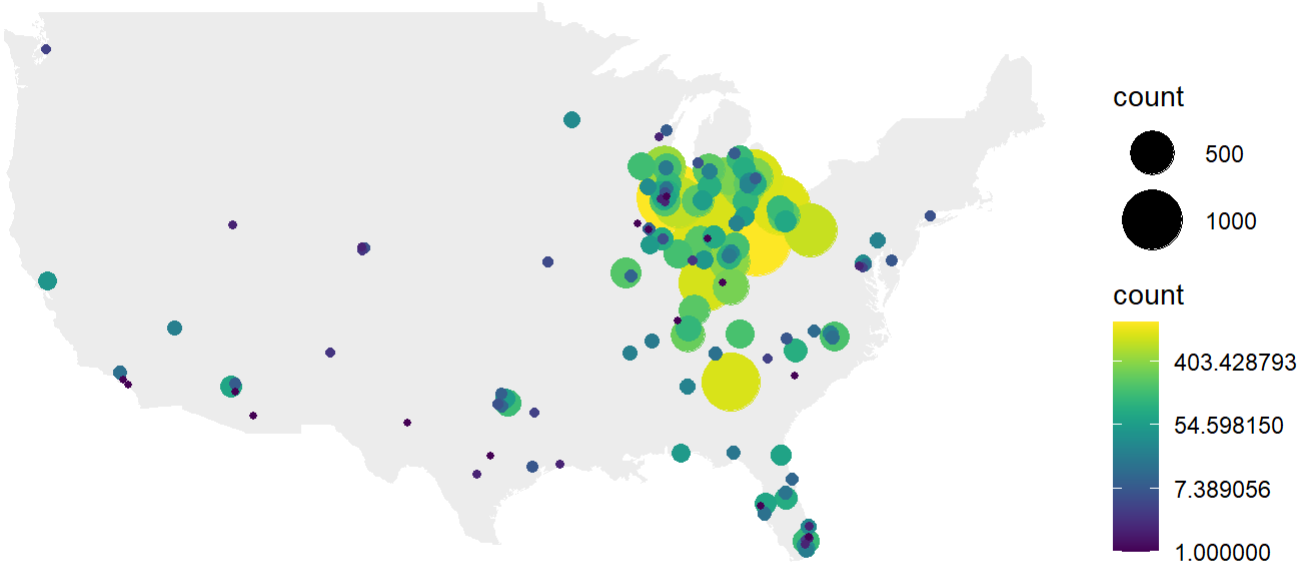
USAsp <- SpatialPolygonsDataFrame(USApolygons, data = dat2)

#plot it: spplot(USAsp['count'])

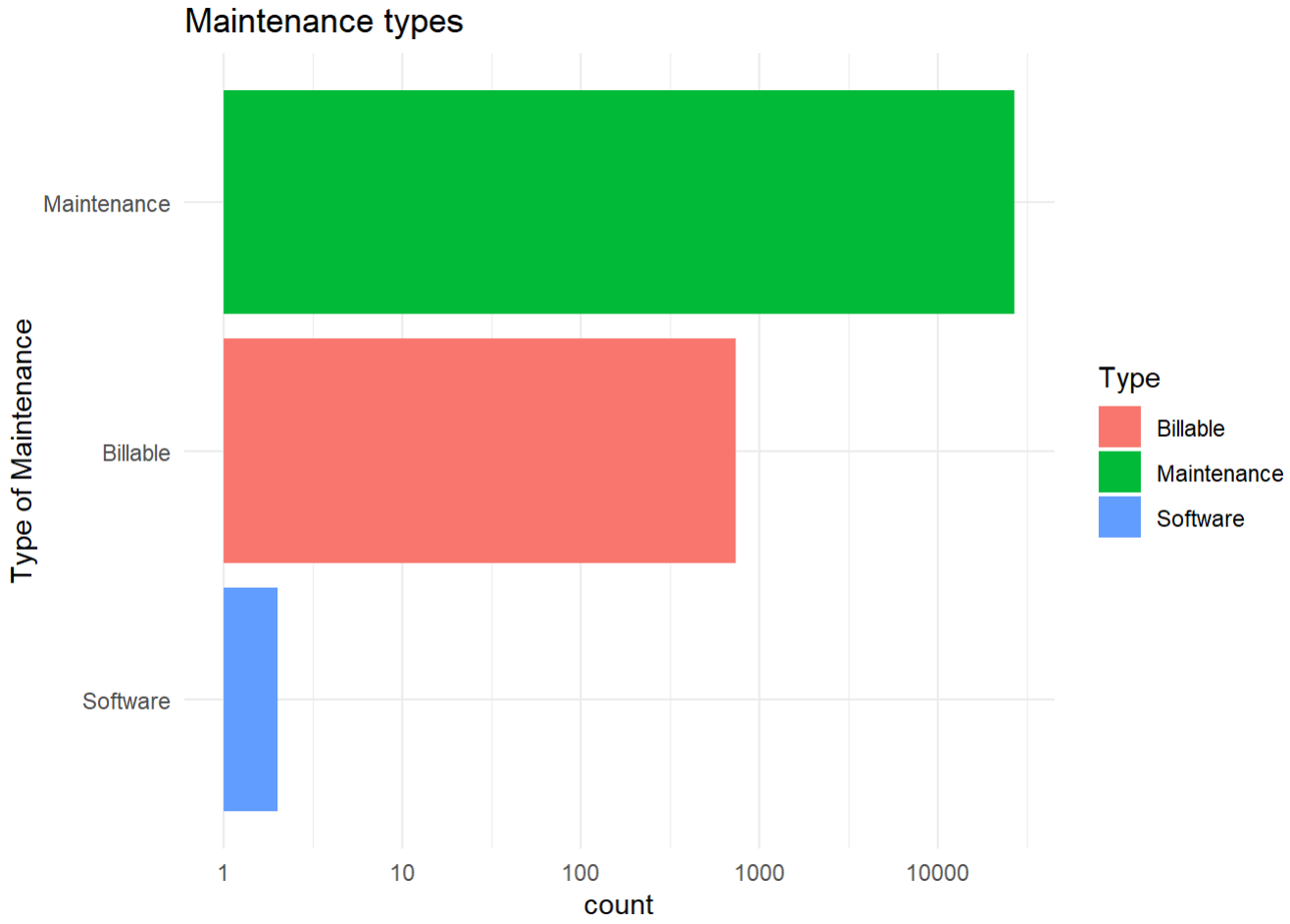
Map of ticket calls by city




count
500
1000
count
403.428793
54.598150
7.389056
1.000000

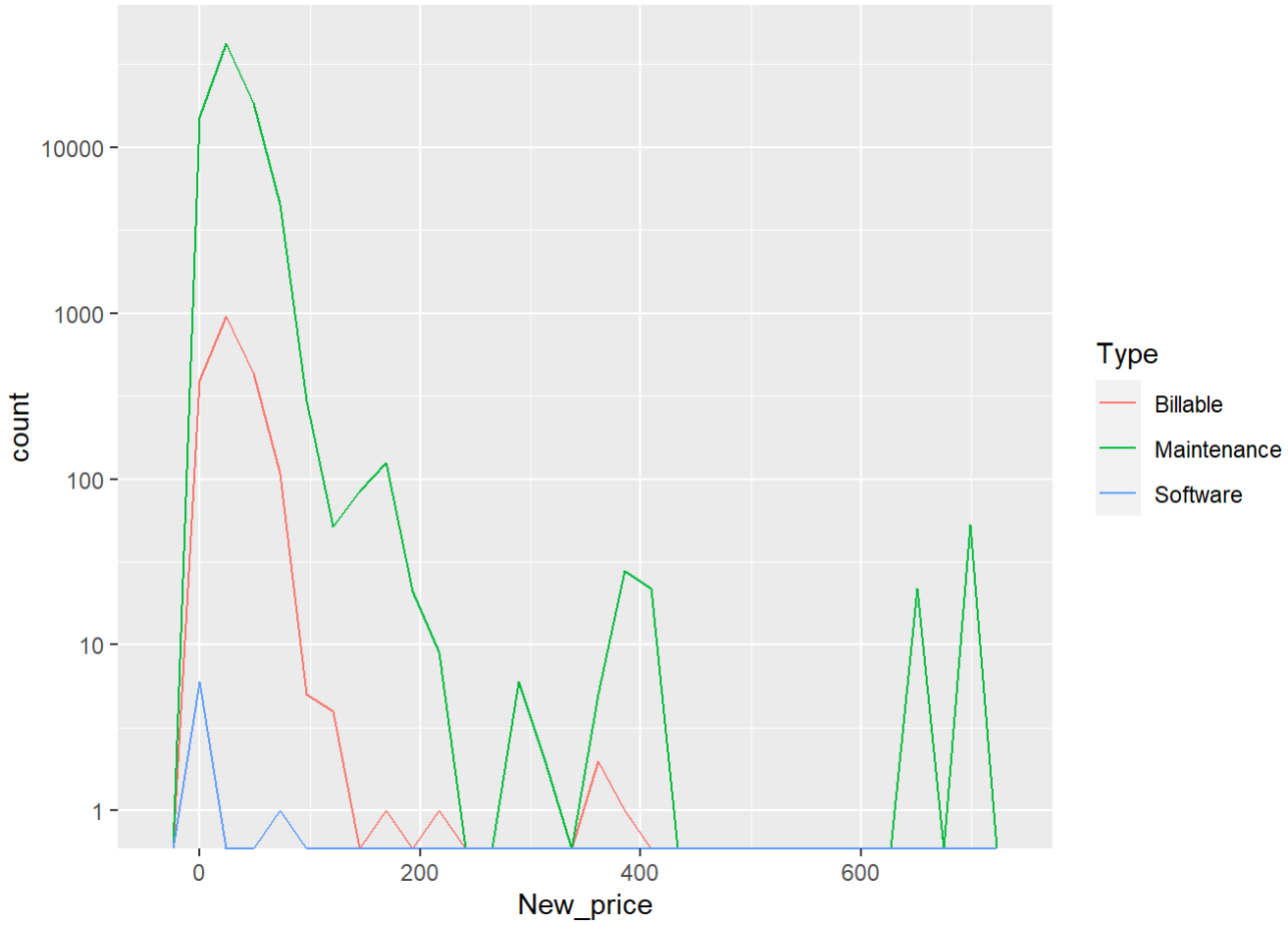

count
500
1000
count
403.428793
54.598150
7.389056
1.000000

Type of maintenance: The bar plot below shows the type of services the company renders to its customers. Maintenance is the most services the company offers followed by the billable and the software type of service.



Services rendered and their cost: One of the Objective of this project, is to explore the SEI services rendered to our customer and also understand the pricing of equipment parts used. The graph below shows the distribution of prices of equipment parts associated with the kind of services provided. The plots that maintenance is the most service rendered to our customers followed by billage and software respectively. Also I noticed that most of the price of were approaching zero , which implies that a frequency maintenance scheme issued in the initial contract.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

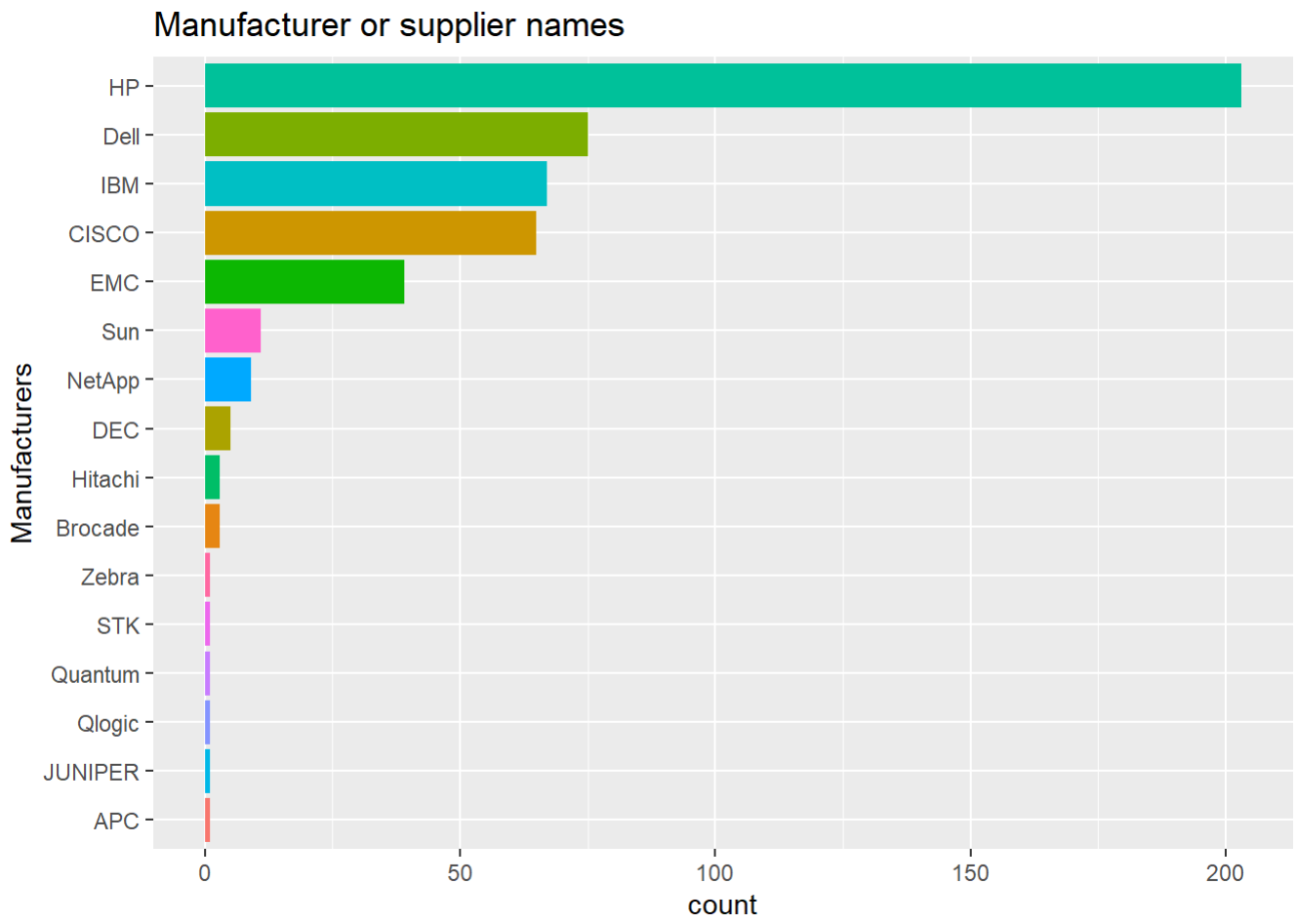


Location: There are a lot of resources needed to help for an effective operation, and I would like to highlight one them which is the equipment model. We would like to know how often this material is used in various state, which you give an idea about how to sock our warehouse at different state. The table below shows the different combination of various model parts frequent usage at different states.

```
## # A tibble: 687 x 3
## # Groups:   ModelNumber [396]
##    ModelNumber       AloState count
##    <chr>             <chr>    <int>
##  1 PROLIANT BL460CG8 IL          17
##  2 PROLIANT DL380PG8 IL           9
##  3 PROLIANT BL460CG7 OH           8
##  4 CP-7925G-A-K9     OH           7
##  5 POWEREDGE 2950    OH           7
##  6 POWEREDGE R710    OH           7
##  7 PROLIANT BL460CG6 GA           7
##  8 PROLIANT BL460CG7 PA           7
##  9 PROLIANT DL360G7  IL           7
## 10 PROLIANT DL380G5  MI           7
## # ... with 677 more rows
```

```
## # A tibble: 211 x 3
## # Groups:   AloState [37]
##    AloState PrimaryEngineer count
##    <chr>    <chr>           <int>
##  1 AL       8680                4
##  2 AL       8716                1
##  3 AZ       8719                3
##  4 AZ       <NA>                2
##  5 AZ       8675                1
##  6 AZ       8745                1
##  7 AZ       8930                1
##  8 CA       8930                3
##  9 CA       8927                2
## 10 CA       8514                1
## # ... with 201 more rows
```

Manufacturing companies: There are a lot of manufacturing companies that we business relationship with to provided the company with quality equipment parts for our operations. The plot below helps us to know the company we deem as priority, since we use a lot of their products. From the graph, it shows that HP is the company that produces a lot of our equipment part for maintenance and services.



Cost of equipment used to service each company This is to find how much does maintaining service cost for each company. Company with AgPKey number 8337 has the highest cost of service at $78,960. The company is in a service industry. The high cost but low frequency could be from the used parts are extremely expensive.

```
## # A tibble: 10 x 3
##    AloPKey.x TotalCost count
##        <dbl>     <dbl> <int>
##  1       302         0    62
##  2       482       660    20
##  3      1396        10     2
##  4      1461         0    52
##  5      1566       625    15
##  6      1590      3200    81
##  7      2196      2280    20
##  8      2692         0    58
##  9      2978         0   505
## 10      3089       361     1
```

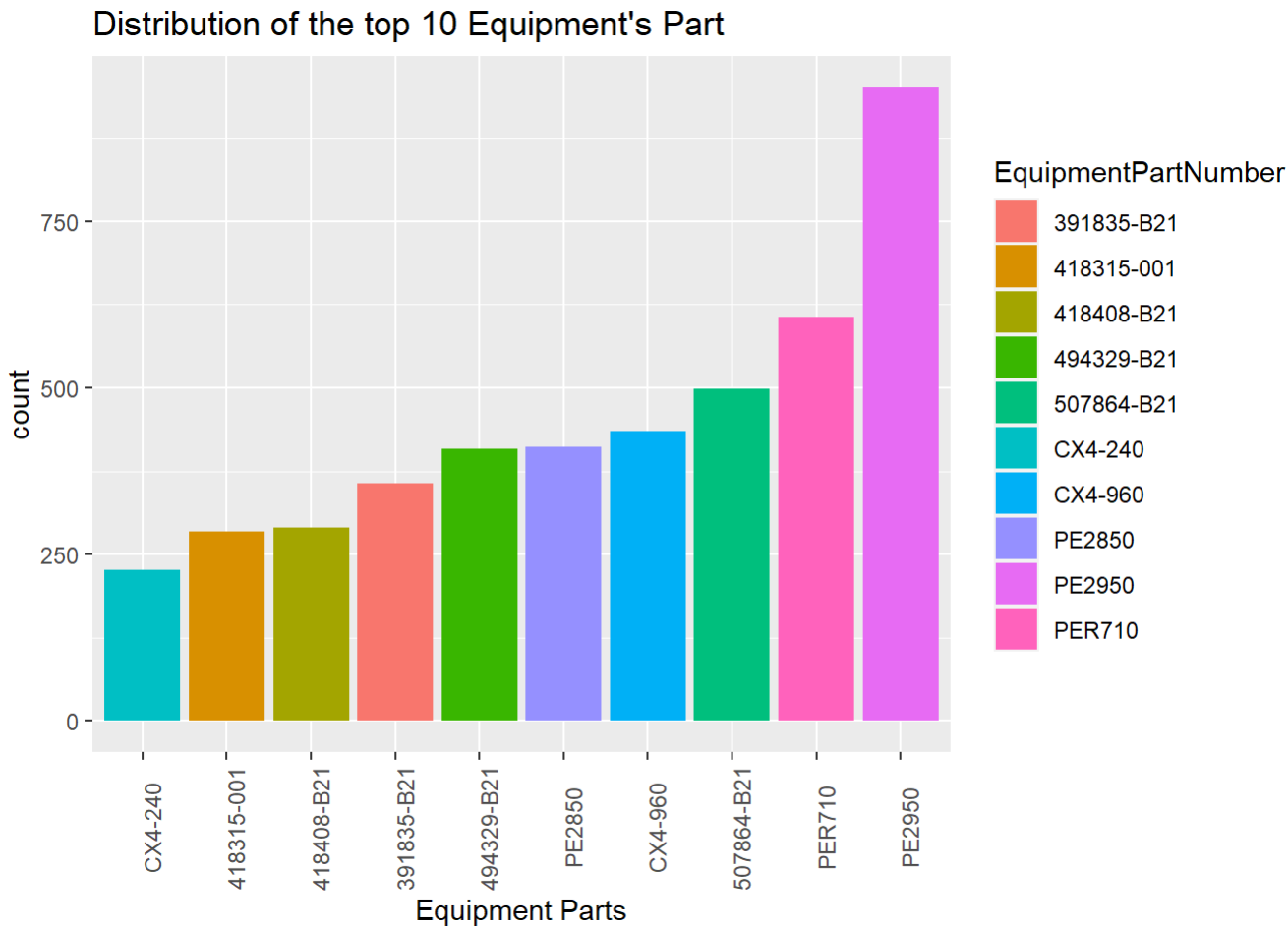
Cost for maintaining by AgrPKey

```
##    AgrPKey TotalCost count CustomerSinceDate ContractStartDate
## 1     8337     78960    24        2010-06-01        2016-06-01
## 2     3862     50347  1829        2005-04-01        2017-04-01
## 3     8011     50234  3068        2010-05-01        2016-05-01
## 4    13936     47550   634        2015-05-01        2015-05-01
## 5     3645     47468   838        2004-10-01        2016-09-01
## 6    16500     46916   634        2016-09-01        2017-09-01
## 7     5461     41882   974        2007-03-01        2016-07-01
## 8     8884     41882  1113        2011-01-01        2017-01-01
## 9    14273     39308   634        2015-08-04        2015-08-04
## 10    3764     38040   708        2005-01-01        2017-01-01
##                   Duration InvoiceFrequency HQState      Industry
## 1     94521600s (~3 years)        Quarterly      IL       Service
## 2     94608000s (~3 years)          Monthly      MI Manufacturing
## 3     94521600s (~3 years)        Quarterly      IL    Wholesaler
## 4     94608000s (~3 years)           Annual      NC        Unkown
## 5     94521600s (~3 years)          Monthly      IL     Education
## 6     94608000s (~3 years)          Monthly      FL        Unkown
## 7     94521600s (~3 years)        Quarterly      OH   Health Care
## 8    31449600s (~52 weeks)           Annual      KY   Health Care
## 9  94348800s (~2.99 years)           Annual      TN        Unkown
## 10    94521600s (~3 years)        Quarterly      IL       Finance
```

Average cost by service

```
## [1] 19.76665
```

Distribution of the top 10 Equipment's Part It is good to know what parts are important and used the most to be able to be prepared for future events. The graph below shows the equipment parts that has been frequently used the most for the services tickets.



Conclusion

This project is to make exploratory analysis about customer information, SEI service and their equipment maintenance. This project reveals details of SEI services and some of the major objectives of the SEI services is revenue growth, and excellent customer service. SEI could attract more customers in new areas and new industries, and monitor the cost of maintenance and parts. Moreover, SEI could use these insights to plan its business strategy, and the company growth and development.