

Job Openings in NYC (Exploratory Analysis)

```
require(ggplot2)

## Loading required package: ggplot2

require(tidyverse)

## Loading required package: tidyverse

## -- Attaching packages ----- tidyverse
1.2.1 --

## v tibble 1.4.2      v purrr 0.2.5
## v tidyr 0.8.2      v dplyr 0.7.7
## v readr 1.1.1      v stringr 1.3.1
## v tibble 1.4.2      v forcats 0.3.0

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

require(lattice)

## Loading required package: lattice

nyc_data = read_csv('NYC_Jobs.csv')

## Parsed with column specification:
## cols(
##   `Work Location` = col_character(),
##   IT_Salary_From = col_double(),
##   IT_Salary_To = col_double(),
##   NonIT_Salary_from = col_double(),
##   NonIT_Salary_To = col_double(),
##   Annual_salary_from = col_integer(),
##   Annual_Salary_to = col_double(),
##   Daily_Salary_from = col_integer(),
##   Daily_Salary_to = col_integer(),
##   Hourly_Salary_from = col_integer(),
##   Hourly_Salary_to = col_integer(),
##   Annual_Salary_freq = col_integer(),
##   Daily_salary_freq = col_integer(),
##   Hourly_salary_freq = col_integer(),
##   Total_Opening = col_integer(),
##   Non_IT = col_integer(),
##   IT = col_integer(),
##   Full_Time = col_integer(),
##   Part_Time = col_integer()
## )
```

```
nyc_data

## # A tibble: 217 x 19
##   `Work Location` IT_Salary_From IT_Salary_To NonIT_Salary_fr~
##   <chr>          <dbl>          <dbl>          <dbl>
## 1 S.I            0            0            54141
## 2 Manhattan      73148.       90358.       58897.
## 3 Manhattan      0            0            36546.
## 4 Queens          0            0            43246.
## 5 Bronx           0            0            26156.
## 6 Brooklyn       0            0            33875
## 7 S.I            0            0            48492.
## 8 Manhattan      39841        52045        50322.
## 9 Manhattan      0            0            27276.
## 10 Manhattan     56646.       85387.       46892.
## # ... with 207 more rows, and 15 more variables: NonIT_Salary_To <dbl>,
## #   Annual_salary_from <int>, Annual_Salary_to <dbl>,
## #   Daily_Salary_from <int>, Daily_Salary_to <int>,
## #   Hourly_Salary_from <int>, Hourly_Salary_to <int>,
## #   Annual_Salary_freq <int>, Daily_salary_freq <int>,
## #   Hourly_salary_freq <int>, Total_Opening <int>, Non_IT <int>, IT <int>,
## #   Full_Time <int>, Part_Time <int>
```

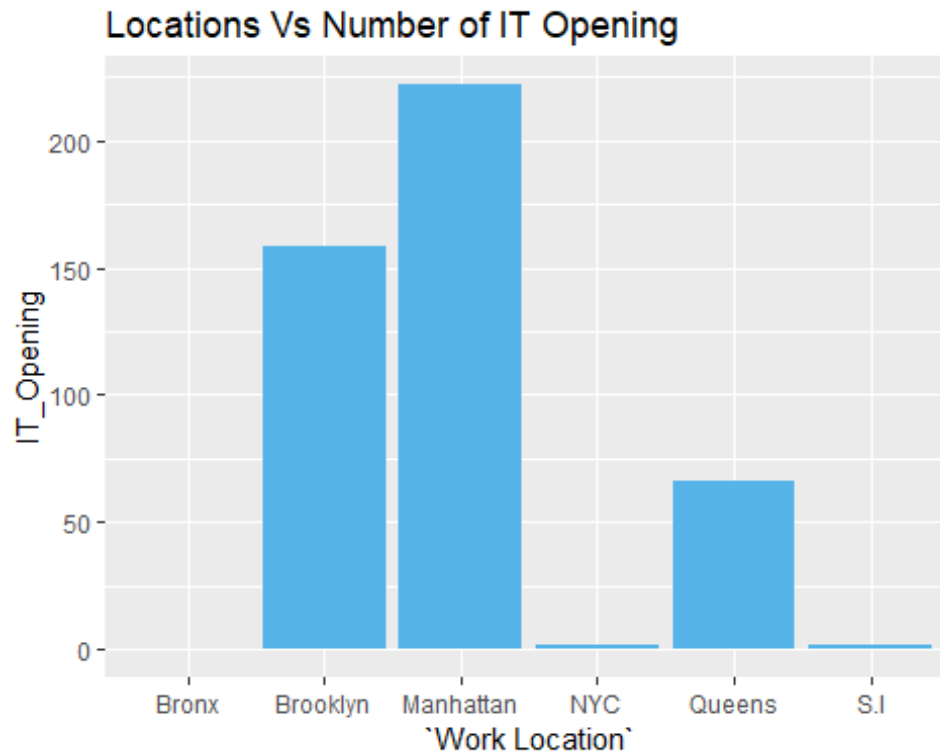
1. Relationship between Work Location and IT and Non IT Job Opening with graph and explanation.

Ans.1

```
IT_Location = nyc_data %>% group_by(`Work Location`) %>% summarize(IT_Opening
= sum(IT))
IT_Location

## # A tibble: 6 x 2
##   `Work Location` IT_Opening
##   <chr>          <int>
## 1 Bronx            0
## 2 Brooklyn        158
## 3 Manhattan        222
## 4 NYC              2
## 5 Queens           66
## 6 S.I              2

ggplot(data=IT_Location, aes(x=`Work Location`, y=IT_Opening)) +
  geom_bar(stat="identity", position=position_dodge(), fill="#56B4E9")+
  labs(title="Locations Vs Number of IT Opening")
```



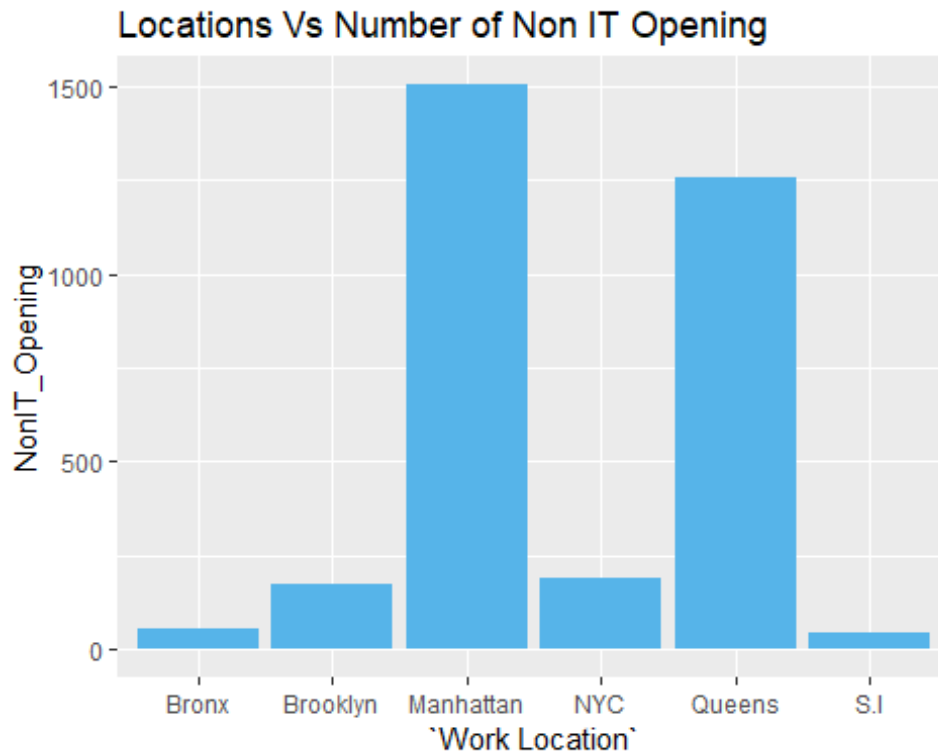
Above graph shows the relationship between Work Location and IT Job Openings.

In NYC, in Manhattan has the maximum number of IT Job Opening around 250. After Manhattan, there are around 160 IT Job opening in Brooklyn. Whereas in the Bronx, there are not any IT Job Opening.

```
NonIT_Location = nyc_data %>% group_by(`Work Location`) %>%
  summarize(NonIT_Opening = sum(Non_IT))
NonIT_Location

## # A tibble: 6 x 2
##   `Work Location` NonIT_Opening
##   <chr>           <int>
## 1 Bronx             56
## 2 Brooklyn          174
## 3 Manhattan        1504
## 4 NYC               187
## 5 Queens           1254
## 6 S.I               43

ggplot(data=NonIT_Location, aes(x=`Work Location`, y=NonIT_Opening)) +
  geom_bar(stat="identity", position=position_dodge(), fill="#56B4E9") +
  labs(title="Locations Vs Number of Non IT Opening")
```



Above graph shows the relationship between Work Location and Non-IT Job Openings.

In NYC, there are maximum non-IT Job Openings in Manhattan around 1500. After Manhattan, in Queens, there are Non-IT Job Openings around 1250. Whereas in S.I(Staten Island), there is almost no job opening in Non-IT.

2. Relationship between Work Location and salary range according to IT and NonIT Opening.

Ans.2

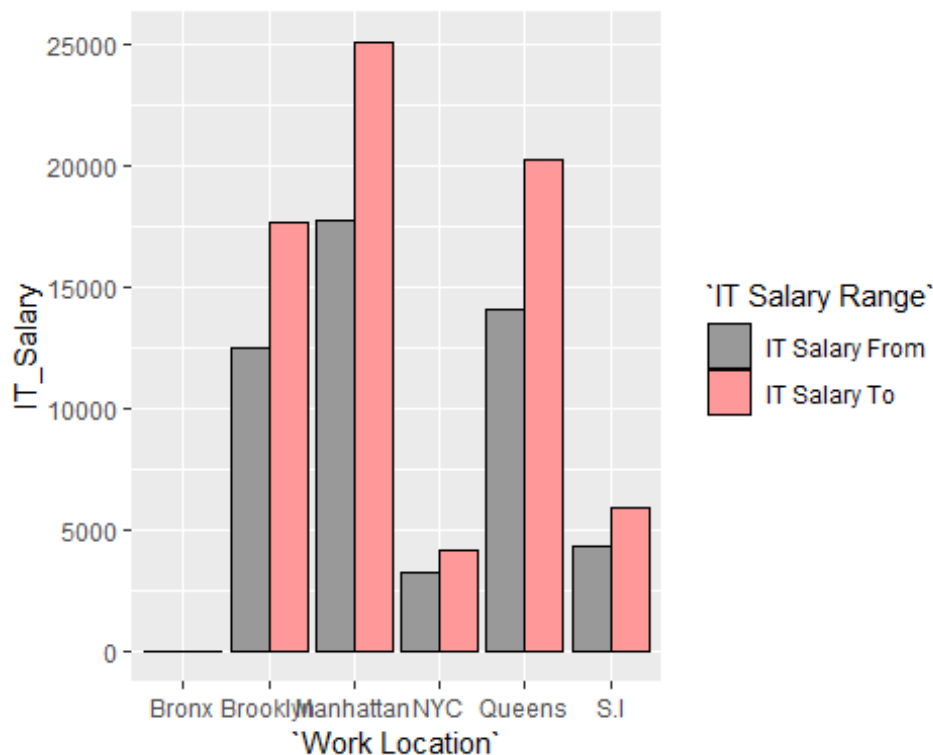
```
##>% filter(IT != 0)
IT_salary_range = nyc_data %>% group_by(`Work Location`) %>% summarize('IT
Salary From' = round(mean(IT_Salary_From)), 'IT Salary To' =
round(mean(IT_Salary_To)))
IT_salary_range

## # A tibble: 6 x 3
##   `Work Location` `IT Salary From` `IT Salary To`
##   <chr>          <dbl>          <dbl>
## 1 Bronx          0              0
## 2 Brooklyn      12439          17617
## 3 Manhattan     17714          25056
## 4 NYC           3215           4142
## 5 Queens       14030          20236
## 6 S.I           4336           5909

IT_Sal = gather(IT_salary_range, 'IT Salary Range', IT_Salary, 2:3)
IT_Sal
```

```
## # A tibble: 12 x 3
##   `Work Location` `IT Salary Range` IT_Salary
##   <chr>          <chr>          <dbl>
## 1 Bronx          IT Salary From          0
## 2 Brooklyn       IT Salary From       12439
## 3 Manhattan       IT Salary From       17714
## 4 NYC             IT Salary From        3215
## 5 Queens          IT Salary From       14030
## 6 S.I             IT Salary From        4336
## 7 Bronx          IT Salary To           0
## 8 Brooklyn       IT Salary To       17617
## 9 Manhattan       IT Salary To       25056
## 10 NYC            IT Salary To        4142
## 11 Queens         IT Salary To       20236
## 12 S.I            IT Salary To        5909

ggplot(data=IT_Sal, aes(x=`Work Location`, y=IT_Salary, fill=`IT Salary
Range`)) +
  geom_bar(stat="identity", position=position_dodge(), colour="black") +
  scale_fill_manual(values=c("#999999", "#ff9999"))
```



Above graph shows the relationship between Work Location and Salary Range of IT job openings. Brown bar show minimum salary and Yellow bar shows maximum salary for a particular IT Job Opening. Note, there are no It Job Openings in Bronx so, we will not consider that. We can see that Manhattan has maximum salary range which is from 17500 to 25000 and this is average salary range. S.I has minimum salary range which is from around 4000 to 7000.

```

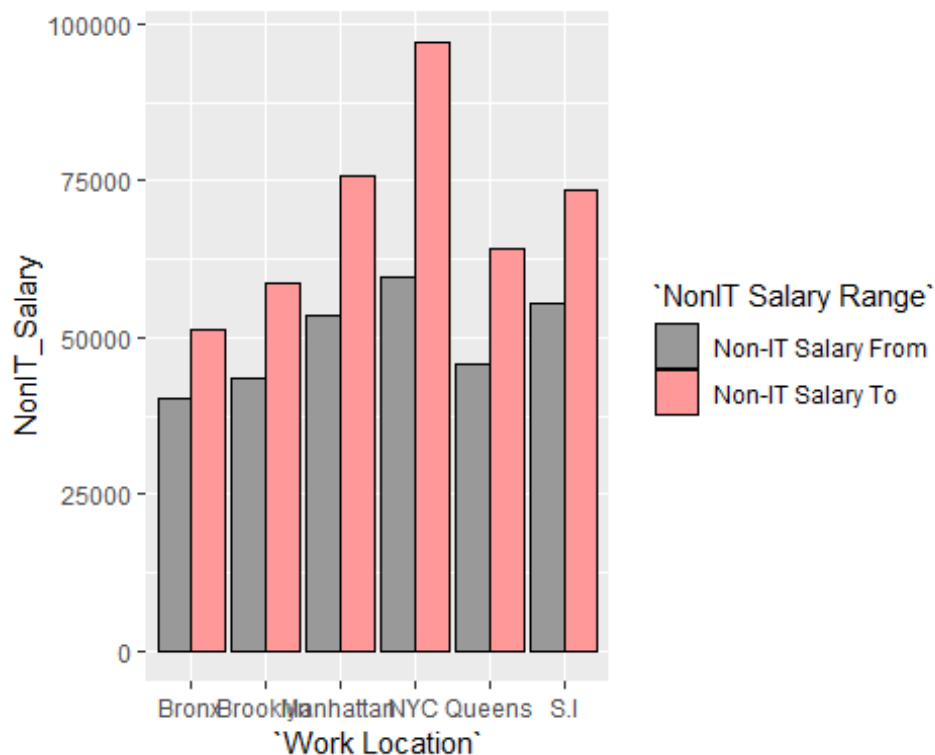
NonIT_salary_range = nyc_data %>% group_by(`Work Location`) %>%
  summarize('Non-IT Salary From' = round(mean(NonIT_Salary_from)), 'Non-IT
Salary To' = round(mean(NonIT_Salary_To)))

NonIT_Sal = gather(NonIT_salary_range, 'NonIT Salary Range', NonIT_Salary,
2:3)
NonIT_Sal

## # A tibble: 12 x 3
##   `Work Location` `NonIT Salary Range` NonIT_Salary
##   <chr>          <chr>          <dbl>
## 1 Bronx          Non-IT Salary From          40310
## 2 Brooklyn       Non-IT Salary From          43599
## 3 Manhattan      Non-IT Salary From          53680
## 4 NYC            Non-IT Salary From          59801
## 5 Queens         Non-IT Salary From          45898
## 6 S.I            Non-IT Salary From          55448
## 7 Bronx          Non-IT Salary To           51371
## 8 Brooklyn       Non-IT Salary To           58639
## 9 Manhattan      Non-IT Salary To           75862
## 10 NYC           Non-IT Salary To           97122
## 11 Queens        Non-IT Salary To           64316
## 12 S.I           Non-IT Salary To           73450

ggplot(data=NonIT_Sal, aes(x=`Work Location`, y=NonIT_Salary, fill=`NonIT
Salary Range`)) +
  geom_bar(stat="identity", position=position_dodge(), colour="black") +
  scale_fill_manual(values=c("#999999", "#ff9999"))

```



Above graph shows the relationship between Work Location and Salary Range of Non-IT job openings. Brown bar show minimum salary and Yellow bar shows maximum salary for a particular IT Job Opening.

We can see that NYC has maximum salary range which is around from 60000 to 95000 and this is average salary range. If we look particular NYC boroughs. Then Manhattan has maximum salary range for Non-IT Job Opening which is around 55000 to 76000. Bronx has minimum salary range which is from around 40000 to 52000. These ranges are average value of salary range.

3. What is the average maximum and minimum annual and hourly salary in each locations.

Ans.3

```
df2 = rename(nyc_data ,AvgMinAnnualSalary = Annual_salary_from,
AvgMaxAnnualSalary= Annual_Salary_to, AvgMinHourlySalary = Hourly_Salary_from
, AvgMaxHourlySalary = Hourly_Salary_to)

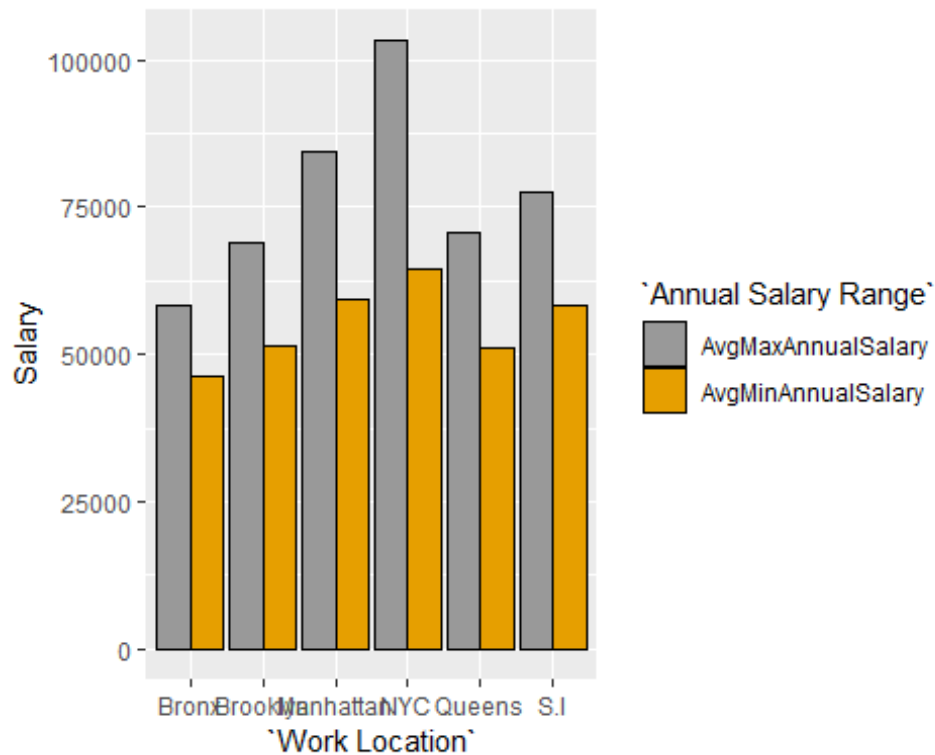
dd = df2 %>% group_by(`Work Location`) %>% summarise(AvgMinAnnualSalary =
round(mean(AvgMinAnnualSalary)),
AvgMaxAnnualSalary=round(mean(AvgMaxAnnualSalary)), AvgMinHourlySalary =
round(mean(AvgMinHourlySalary)) ,
AvgMaxHourlySalary=round(mean(AvgMaxHourlySalary)))

dd

## # A tibble: 6 x 5
##   `Work Location` AvgMinAnnualSal~ AvgMaxAnnualSal~ AvgMinHourlySal~
##   <chr>           <dbl>           <dbl>           <dbl>
## 1 Bronx          46299           58239             6
## 2 Brooklyn      51360           69114             6
## 3 Manhattan     59354           84439             5
## 4 NYC           64532          103325             4
## 5 Queens        51034           70615            10
## 6 S.I           58266           77464             5
## # ... with 1 more variable: AvgMaxHourlySalary <dbl>
```

The above analysis result shows the average minimum and maximum salary in all the locations. The maximum and minimum average salary is annually and hourly based. This means that our analysis collect all the minimum and maximum salaries location wise and then calculates the average of all to give the results.

```
dd3 =gather(dd, 'Annual Salary Range', Salary, 2:3)
ggplot(data=dd3, aes(x=`Work Location`, y=Salary, fill=`Annual Salary
Range`)) +
  geom_bar(stat="identity", position=position_dodge(), colour="black") +
  scale_fill_manual(values=c("#999999", "#E69F00"))
```



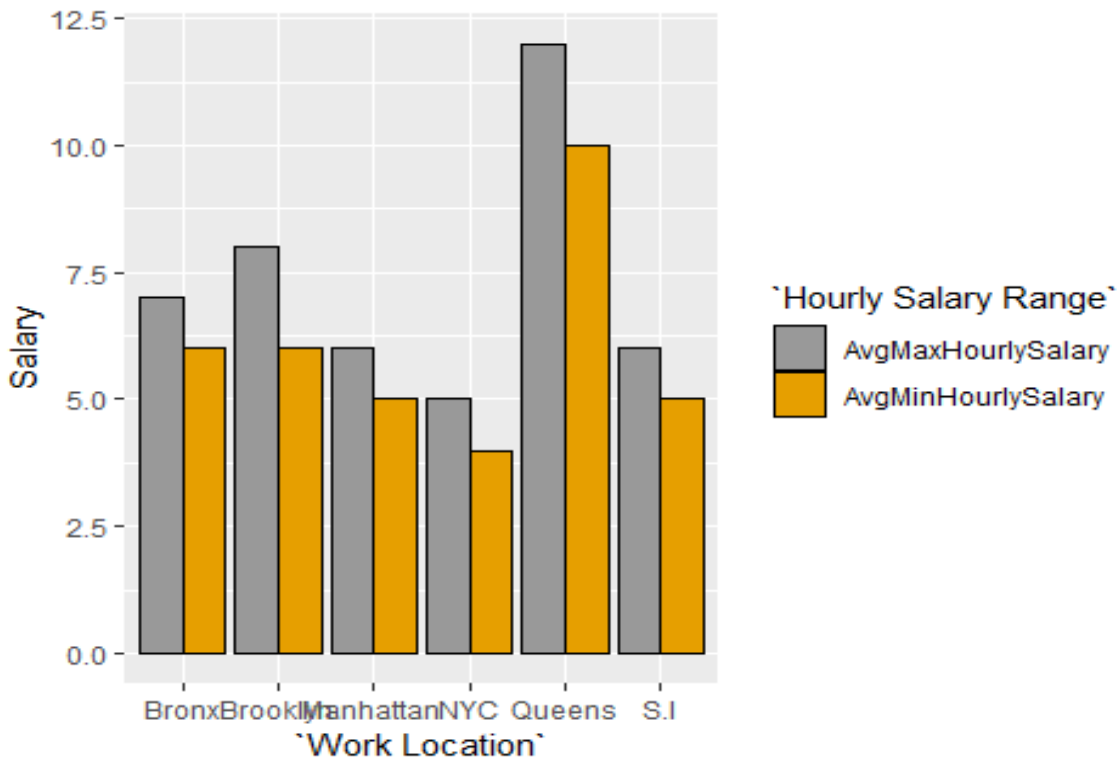
The above plot shows the Annual average maximum and minimum salary on all locations. The average maximum salary is in NYC and least maximum salary is in Bronx. The highest average minimum annual salary is in NYC and least average minimum is at Bronx.

```
hourly =gather(dd, 'Hourly Salary Range', Salary, 4:5)
```

```
hourly
```

```
## # A tibble: 12 x 5
##   `Work Location` AvgMinAnnualSal~ AvgMaxAnnualSal~ `Hourly Salary ~
##   <chr>           <dbl>           <dbl> <chr>
## 1 Bronx          46299          58239 AvgMinHourlySal~
## 2 Brooklyn       51360          69114 AvgMinHourlySal~
## 3 Manhattan      59354          84439 AvgMinHourlySal~
## 4 NYC            64532         103325 AvgMinHourlySal~
## 5 Queens         51034          70615 AvgMinHourlySal~
## 6 S.I            58266          77464 AvgMinHourlySal~
## 7 Bronx          46299          58239 AvgMaxHourlySal~
## 8 Brooklyn       51360          69114 AvgMaxHourlySal~
## 9 Manhattan      59354          84439 AvgMaxHourlySal~
## 10 NYC           64532         103325 AvgMaxHourlySal~
## 11 Queens        51034          70615 AvgMaxHourlySal~
## 12 S.I           58266          77464 AvgMaxHourlySal~
## # ... with 1 more variable: Salary <dbl>
```

```
ggplot(data=hourly, aes(x=`Work Location`, y=Salary, fill=`Hourly Salary
Range`)) +
  geom_bar(stat="identity", position=position_dodge(), colour="black") +
  scale_fill_manual(values=c("#999999", "#E69F00"))
```

The above plot shows that average maximum and minimum hourly salary on all locations. The average maximum salary is in Queens and least maximum salary is in NYC. The highest average minimum hourly salary is in Queens and least average minimum is at NYC.

4. How many part-time and full time jobs are there in each location.

Ans.4

```
df3<- nyc_data %>% group_by(`Work Location`) %>% summarise(Part_Time =
sum(Part_Time), Full_Time=sum(Full_Time))
df3
```

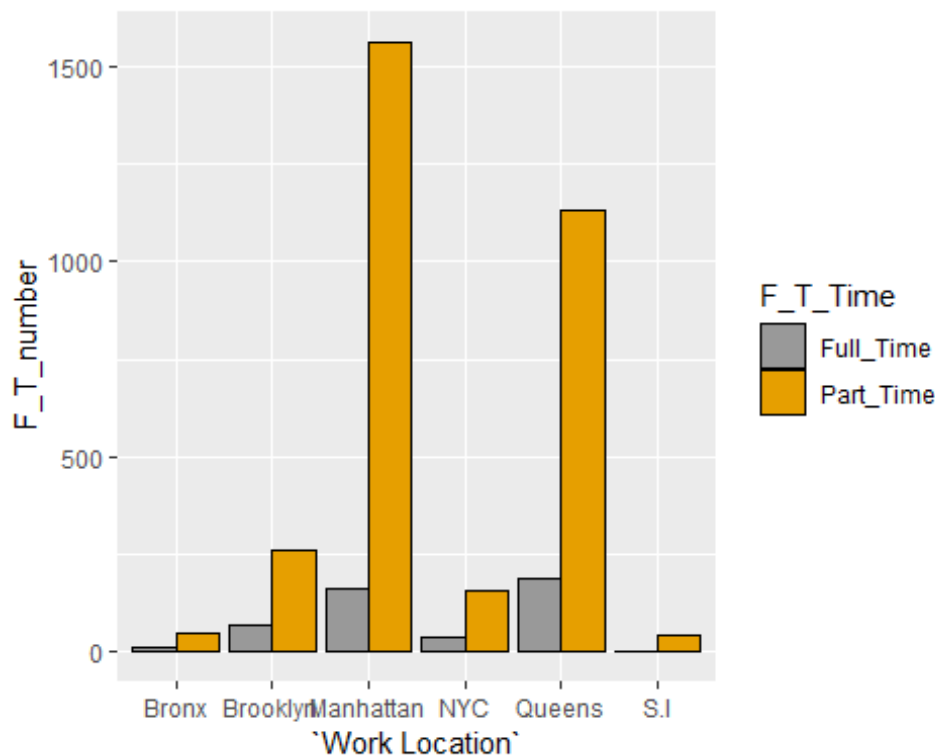
```
## # A tibble: 6 x 3
##   `Work Location` Part_Time Full_Time
##   <chr>           <int>    <int>
## 1 Bronx             46         10
## 2 Brooklyn          261         69
## 3 Manhattan        1564        162
## 4 NYC               155         34
## 5 Queens           1134        188
## 6 S.I               43          2
```

```
df4=gather(df3,F_T_Time,F_T_number,2:3)
df4
```

```
## # A tibble: 12 x 3
##   `Work Location` F_T_Time F_T_number
##   <chr>          <chr>    <int>
## 1 Bronx        Part_Time         46
## 2 Brooklyn    Part_Time        261
## 3 Manhattan    Part_Time       1564
```

```
## 4 NYC Part_Time 155
## 5 Queens Part_Time 1134
## 6 S.I Part_Time 43
## 7 Bronx Full_Time 10
## 8 Brooklyn Full_Time 69
## 9 Manhattan Full_Time 162
## 10 NYC Full_Time 34
## 11 Queens Full_Time 188
## 12 S.I Full_Time 2
```

```
ggplot(data=df4, aes(x=`Work Location`, y=F_T_number, fill=F_T_Time)) +
  geom_bar(stat="identity", position=position_dodge(), colour="black") +
  scale_fill_manual(values=c("#999999", "#E69F00"))
```



The above result describes that 'Manhattan' has the maximum number of Part_Time jobs =1564, where as S.I. has the lowest number of Part_Time jobs=43. On the other hand, Queens has the maximum number of Full_Time jobs=188 , whereas S.I again as the lowest number of Full_Time jobs= 2. The ranking for Part_Time would be - Manhattan,Queens,Brooklyn,NYC,Bronx,S.I. The ranking for Full_Time would be - Queens,Manhattan,Brooklyn,NYC,Bronx,S.I.

- Relationship between salary frequency based on location. Which Work Location has the most jobs opening provided with Annual Salary/Daily Salary/Hourly. Salary?

Ans.5

1.Locations Vs AnnualSalaryFreq

```

jobannual <- aggregate(nyc_data$Annual_Salary_freq, by=list(nyc_data$`Work
Location`), FUN=sum)
jobannual

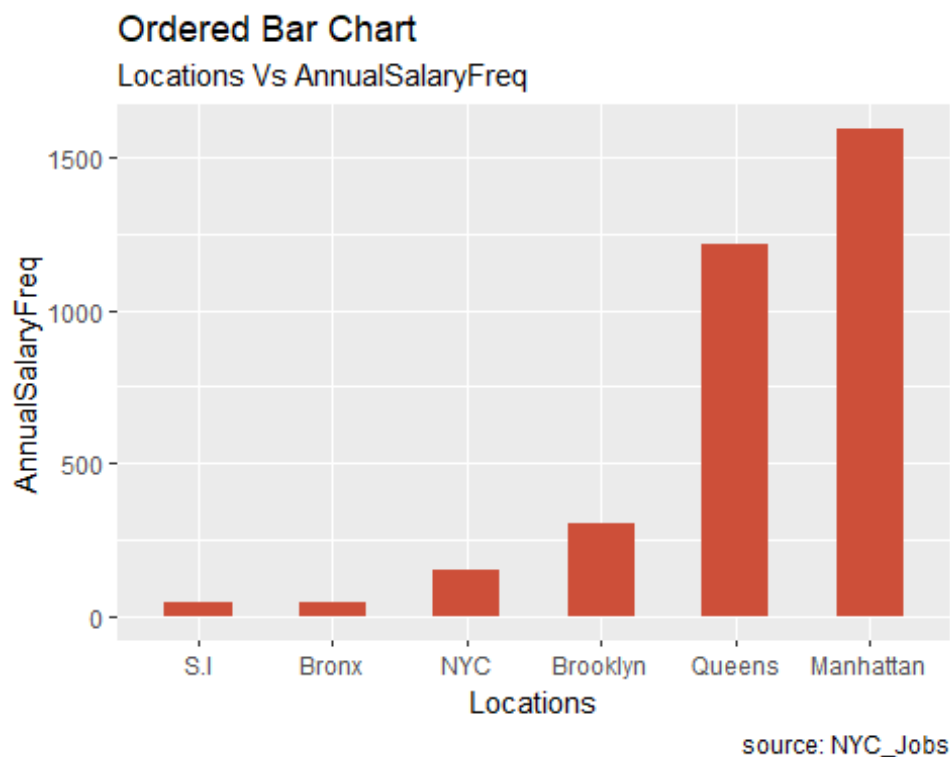
##      Group.1      x
## 1      Bronx      43
## 2    Brooklyn    305
## 3    Manhattan 1592
## 4         NYC     151
## 5      Queens 1215
## 6         S.I      41

colnames(jobannual) <- c("Locations", "AnnualSalaryFreq")
jobannual <- jobannual[order(jobannual$AnnualSalaryFreq), ]
jobannual$Locations <- factor(jobannual$Locations, levels =
jobannual$Locations)
head(jobannual,8)

##   Locations AnnualSalaryFreq
## 6         S.I              41
## 1        Bronx              43
## 4         NYC             151
## 2    Brooklyn             305
## 5      Queens            1215
## 3    Manhattan            1592

# Draw plot
ggplot(jobannual, aes(x=Locations, y=AnnualSalaryFreq)) +
geom_bar(stat="identity", width=.5, fill="tomato3") + labs(title="Ordered Bar
Chart", subtitle="Locations Vs AnnualSalaryFreq", caption="source: NYC_Jobs")

```



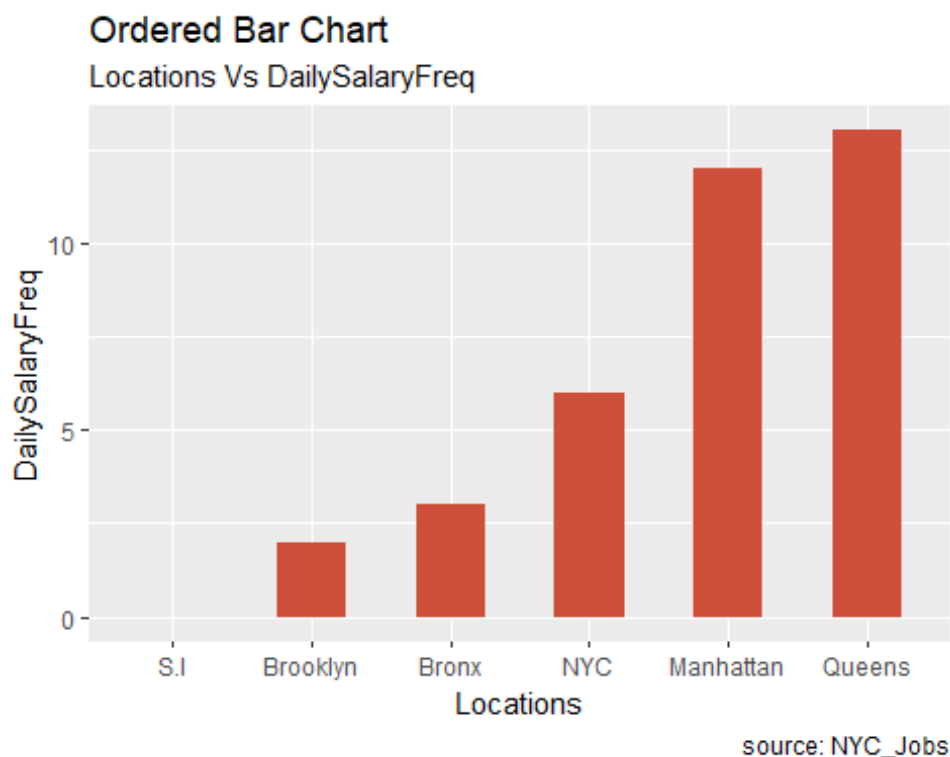
From the above graph `Locations Vs AnnualSalaryFreq`, it can be seen that *Manhattan has the most jobs opening, total of 1592, provide with annual salary*. The ranking will be: Manhattan, Queens, Brooklyn, Bronx, Staten Island.

2. Locations Vs DailySalaryFreq

```
jobdaily <- aggregate(nyc_data$Daily_salary_freq, by=list(nyc_data$`Work
Location`), FUN=sum)
colnames(jobdaily) <- c("Locations", "DailySalaryFreq")
jobdaily <- jobdaily[order(jobdaily$DailySalaryFreq), ]
jobdaily$Locations <- factor(jobdaily$Locations, levels = jobdaily$Locations)
head(jobdaily,8)
```

```
## Locations DailySalaryFreq
## 6 S.I 0
## 2 Brooklyn 2
## 1 Bronx 3
## 4 NYC 6
## 3 Manhattan 12
## 5 Queens 13
```

```
ggplot(jobdaily, aes(x=Locations, y=DailySalaryFreq)) +
geom_bar(stat="identity", width=.5, fill="tomato3") + labs(title="Ordered Bar
Chart", subtitle="Locations Vs DailySalaryFreq", caption="source: NYC_Jobs")
```



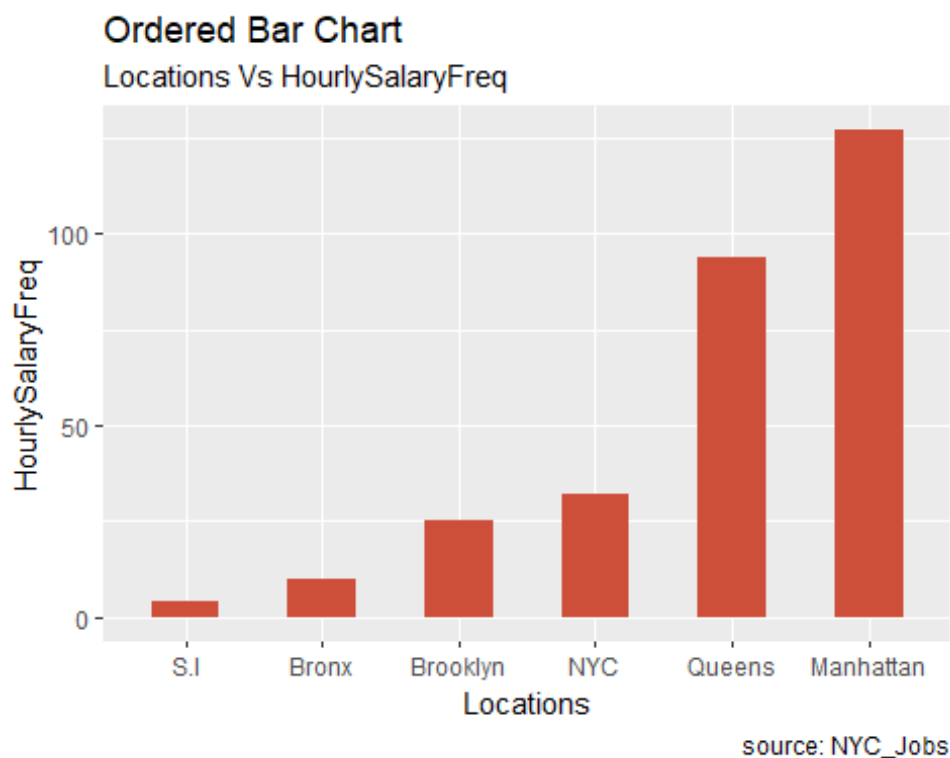
From the above graph `Locations Vs DailySalaryFreq`, it can be seen that *Manhattan has the most jobs opening, total of 13, provide with daily salary*. And there are zero jobs opening provide with daily salary in Staten Island. The ranking will be: Queens, Manhattan, Bronx, Brooklyn, Staten Island.

3.Locations Vs HourlySalaryFreq

```
jobhourly <- aggregate(nyc_data$Hourly_salary_freq, by=list(nyc_data$`Work
Location`), FUN=sum)
colnames(jobhourly) <- c("Locations", "HourlySalaryFreq")
jobhourly <- jobhourly[order(jobhourly$HourlySalaryFreq), ]
jobhourly$Locations <- factor(jobhourly$Locations, levels =
jobhourly$Locations)
head(jobhourly,8)

##  Locations HourlySalaryFreq
## 6      S.I                4
## 1     Bronx               10
## 2  Brooklyn              25
## 4      NYC               32
## 5    Queens              94
## 3  Manhattan            127

ggplot(jobhourly, aes(x=Locations, y=HourlySalaryFreq)) +
geom_bar(stat="identity", width=.5, fill="tomato3") + labs(title="Ordered Bar
Chart", subtitle="Locations Vs HourlySalaryFreq", caption="source: NYC_Jobs")
```



From the above graph “Locations Vs HourlySalaryFreq”, it can be seen that *Manhattan has the most jobs opening, total of 127, provide with Hourly salary. And there are only 4 jobs opening provide with hourly salary in Staten Island.* The ranking will be: Manhattan, Queens, Brooklyn, Bronx, Staten Island.

Observation From the above result it is seen that Manhattan has the most jobs opening on both Annuanl and Hourly salary. Queens has the most jobs opening for Daily Salary and has

the second most jobs opening with Annual and Hourly salary. Staten Island has the least number of jobs opening on all annual, daily, and hourly salary.