

# Factors affect satisfaction of employees

A Capstone project:

Tonia Chu

Under the mentorship:

Dr. Marko Mitic (Data Scientist at Telenor, Belgrade, Serbia)

For the course:

Foundations of Data Science (Springboard)

## I. INTRODUCTION TO THE PROBLEM

In 2015, more than 400,000 employees participated in the Federal Employee Viewpoint Survey (FEVS). The survey includes questions about satisfaction, leadership, and work schedules. The analysis try to identify the working status and give result on in which respect and how to improve the feelings of federal employees. From the analysis of the survey, we can get answers to many interesting questions.

The Global Satisfaction Index measures employee satisfaction about four aspects related to their work: their job, their pay, their organization, and whether they would recommend their organization as a good place to work. Understanding employee satisfaction along these four dimensions can help reduce costs in the long run. Satisfied employees are more likely to stay in their jobs, reducing turnover.

The Global Satisfaction Index is comprised of the following survey items:

I recommend my organization as a good place to work. (Q. 40)

Considering everything, how satisfied are you with your job? (Q. 69)

Considering everything, how satisfied are you with your pay? (Q. 70)

Considering everything, how satisfied are you with your organization? (Q. 71)

**In this study, I want to solve following problems:**

1. Data wrangling, eliminate extra information in the dataset.
2. See the relationship among columns in the data set: does the age has any influence on the answers from questions? And what about gender?
3. The satisfaction of male vs female employees? Plot the relationship between gender (M, F) on answers from questions.
4. The satisfaction of different age groups? Plot the relationship among age groups.

5. Does agency have any influence on answers? Plot the relationship among agencies.

## II. DEEPER DIVE INTO THE DATA SET

After loading the dataset of 2015 survey csv file into R, we can see that there are 421,748 records of 99 columns. Some of the data are NAs and some characters are empty. The 99 columns include 84 survey questions and other demographics items such as age group, gender, supervisory status and so on.

```
> names(survey_all)
[1] "POSTWT" "agency" "PLEVEL1" "PLEVEL2" "Q1"
[6] "Q2" "Q3" "Q4" "Q5" "Q6"
[11] "Q7" "Q8" "Q9" "Q10" "Q11"
[16] "Q12" "Q13" "Q14" "Q15" "Q16"
[21] "Q17" "Q18" "Q19" "Q20" "Q21"
[26] "Q22" "Q23" "Q24" "Q25" "Q26"
[31] "Q27" "Q28" "Q29" "Q30" "Q31"
[36] "Q32" "Q33" "Q34" "Q35" "Q36"
[41] "Q37" "Q38" "Q39" "Q40" "Q41"
[46] "Q42" "Q43" "Q44" "Q45" "Q46"
[51] "Q47" "Q48" "Q49" "Q50" "Q51"
[56] "Q52" "Q53" "Q54" "Q55" "Q56"
[61] "Q57" "Q58" "Q59" "Q60" "Q61"
[66] "Q62" "Q63" "Q64" "Q65" "Q66"
[71] "Q67" "Q68" "Q69" "Q70" "Q71"
[76] "Q72" "Q73" "Q74" "Q75" "Q76"
[81] "Q77" "Q78" "Q79" "Q80" "Q81"
[86] "Q82" "Q83" "Q84" "DSUPER" "DSEX"
[91] "DMINORITY" "DEDUC" "DFEDTEN" "DRETIRE" "DDIS"
[96] "DAGEGRP" "DMIL" "DLEAVING" "RANDOM"
```

First, we select 48 useful columns for analysis. These questions are related with Engagement, Global Satisfaction and the New IQ.

Employee engagement is the employee's sense of purpose. It is evident in their display of dedication, persistence, and effort in their work or overall commitment to their organization and its mission. An agency that engages its employees ensures a work environment where each employee can reach his or her potential, while contributing to the success of the agency. Individual agency performance contributes to success for the entire Federal Government. The Engagement index is made up of three sub-factors:

Leaders Lead, Supervisors, and Intrinsic Work Experience. Each sub-factor reflects a different aspect of an engaging work environment.

The Global Satisfaction Index measures employee satisfaction about four aspects related to their work: their job, their pay, their organization, and whether they would recommend their organization as a good place to work. Understanding employee satisfaction along these four dimensions can help reduce costs in the long run. Satisfied employees are more likely to stay in their jobs, reducing turnover. The effects of turnover are costly, not only in recruitment and on-boarding processes, but also in terms of lost productivity and lower customer satisfaction.

The New IQ identifies behaviors that help create an inclusive environment and is built on the concept that repetition of inclusive behaviors will create positive habits among team members and managers. The New IQ Index indicates the degree to which an environment is inclusive. The New IQ is comprised of the following sub-factors and items:

Fair: Are all employees treated equitably? (Q 23, 24, 25, 37, & 38)

Open: Does management support diversity in all ways? (Q 32, 34, 45, & 55)

Cooperative: Does management encourage communication and collaboration? (Q 58 & 59)

Supportive: Do supervisors value employees? (Q 42, 46, 48, 49, & 50)

Empowering: Do employees have the resources and support needed to excel? (Q 2, 3, 11, & 30)

```
> names(survey)
```

```
[1] "POSTWT" "agency" "Q2" "Q3" "Q4"
[6] "Q6" "Q11" "Q12" "Q23" "Q24"
[11] "Q25" "Q30" "Q32" "Q34" "Q37"
[16] "Q38" "Q40" "Q42" "Q45" "Q46"
[21] "Q47" "Q48" "Q49" "Q50" "Q51"
[26] "Q52" "Q53" "Q54" "Q55" "Q56"
[31] "Q58" "Q59" "Q60" "Q61" "Q69"
[36] "Q70" "Q71" "DSUPER" "DSEX" "DMINORITY"
[41] "DEDUC" "DFEDTEN" "DRETIRE" "DDIS" "DAGEGRP"
[46] "DMIL" "DLEAVING" "RANDOM"
```

Second, we removed rows with NA values. Data set reduced to 300,801 records of 48 columns. Then we randomly select 20k rows of dataset for analysis. All the research and analysis about gender and age group are based on this dataset.

Next, we choose 6 agencies to do the research. We removed the data of other agencies and randomly select 20k rows of dataset for analysis.

### III. EXPLORATION AND INITIAL FINDINGS

Then, we did some analysis to find the relationship of Global Satisfaction with gender, age group and agency.

We use the answers of Q40, Q69, Q70, Q71 to indicate global satisfaction. The questions are as follows:

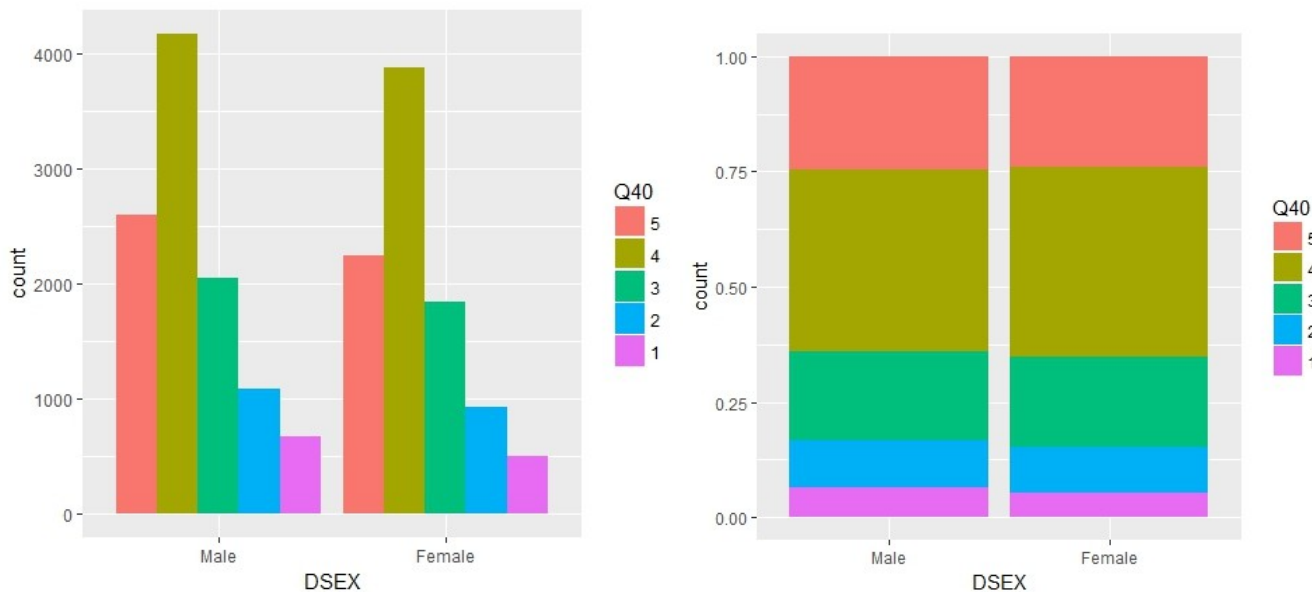
I recommend my organization as a good place to work. (Q40)

Considering everything, how satisfied are you with your job? (Q69)

Considering everything, how satisfied are you with your pay? (Q70)

Considering everything, how satisfied are you with your organization? (Q71)

#### Sex

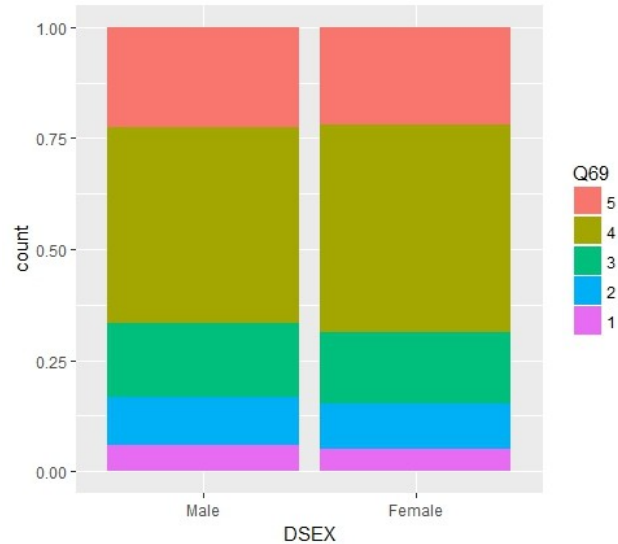
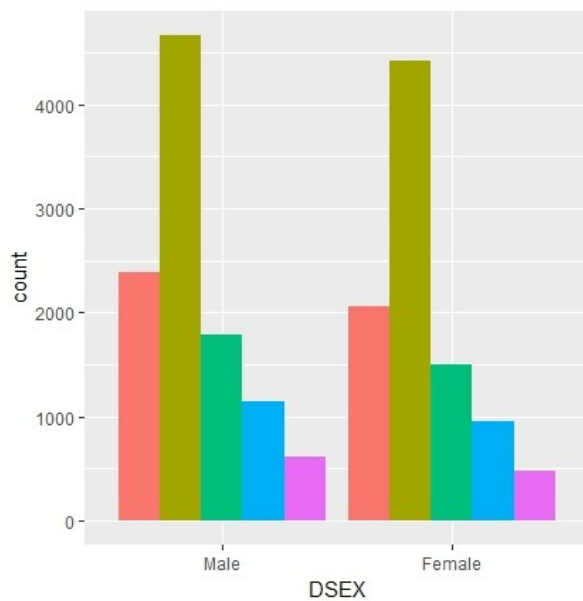


From the plots, we can see that there are no much differences between the Q40 answers of Male vs Female. Females are slightly more satisfied than Males. We can get the same result from the table below:

```
round(prop.table(tab_sex, 1)*100,2)
```

	5	4	3	2	1
Male	24.53	39.43	19.34	10.30	6.40
Female	23.86	41.25	19.65	9.94	5.29

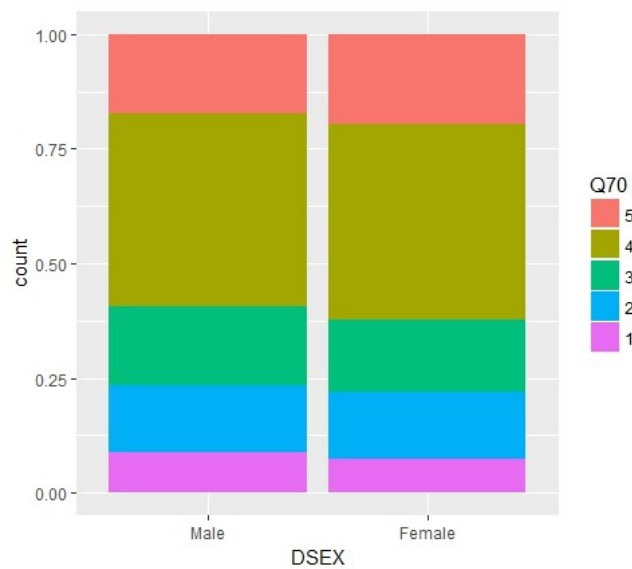
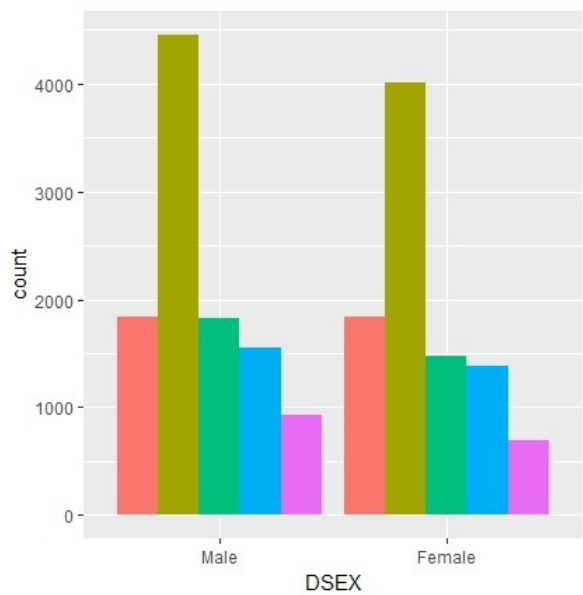
1) Job



```
round(prop.table(tab_sex_Q69, 1)*100,2)
```

	5	4	3	2	1
Male	22.53	44.01	16.88	10.78	5.80
Female	21.85	46.95	15.95	10.20	5.04

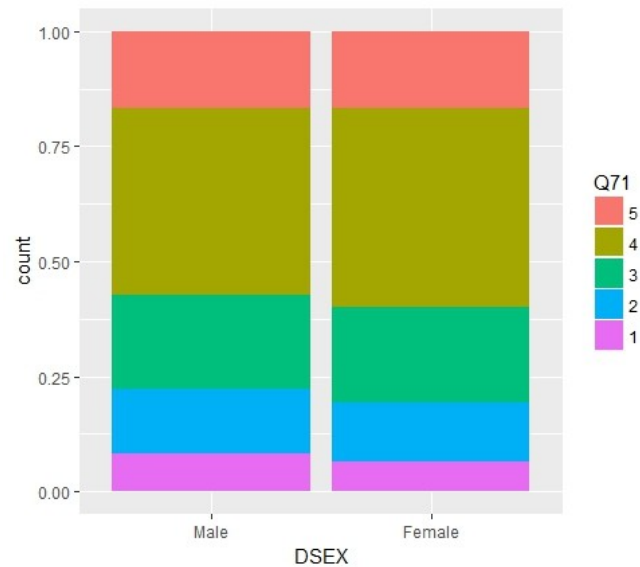
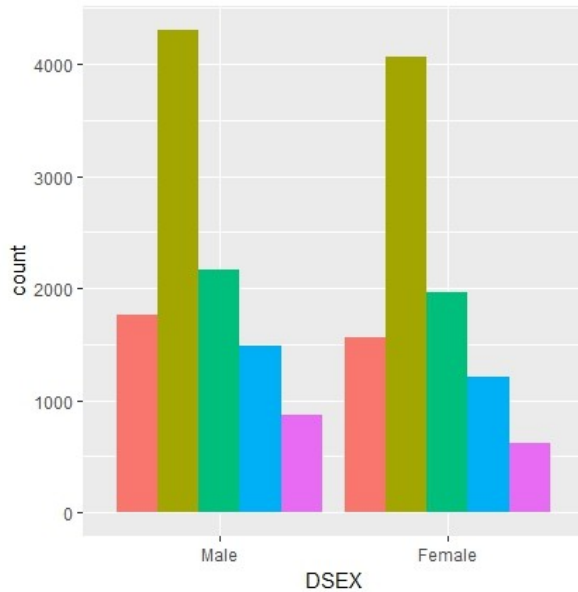
## 2) Pay



```
round(prop.table(tab_sex_Q70, 1)*100,2)
```

	5	4	3	2	1
Male	17.31	42.05	17.26	14.65	8.72
Female	19.52	42.70	15.69	14.72	7.38

## 3) Organization

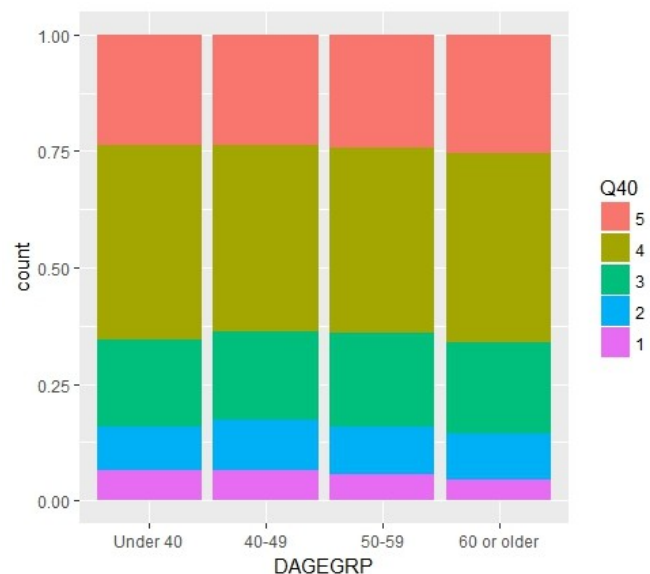
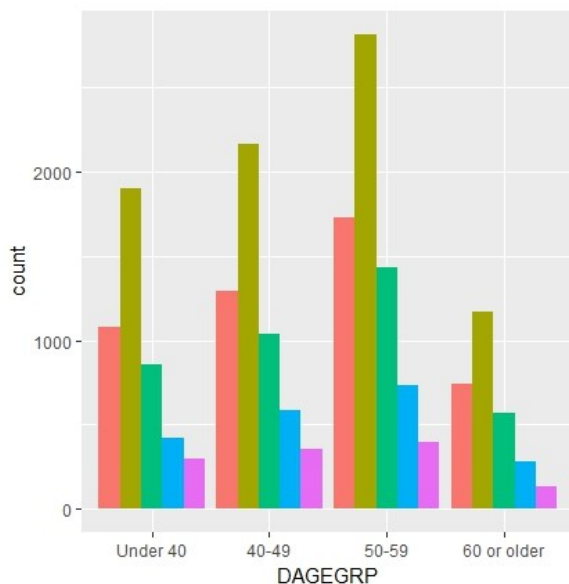


```
round(prop.table(tab_sex_Q71, 1)*100,2)
```

	5	4	3	2	1
Male	16.64	40.63	20.48	13.99	8.25
Female	16.63	43.17	20.83	12.85	6.52

From the plots, we can see that there are no much differences between the answers of Male vs Female. Females are a little more satisfied with job, pay and organization than Males.

## Age group



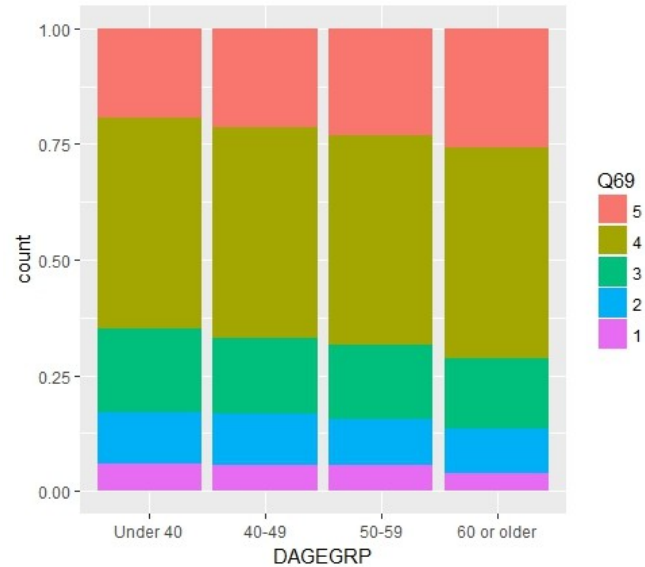
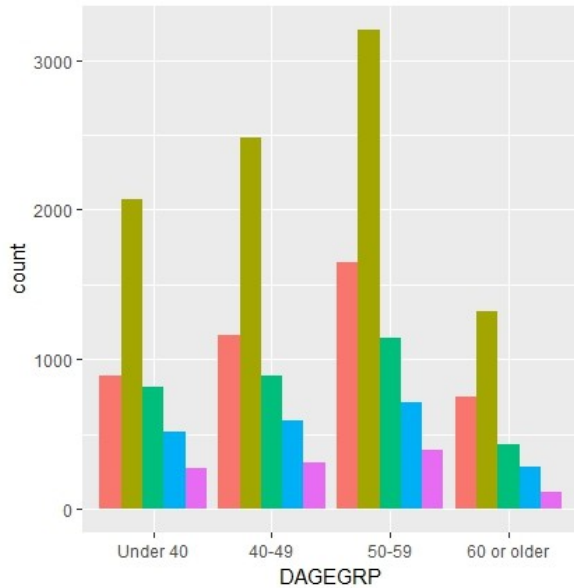
From the plots, we can see that there are slightly differences among the Q40 answers of age groups. Under 40 and 60 or older groups are more

satisfied than 40-49 and 50-59 groups. We can get the same result from the table below:

```
round(prop.table(tab_age, 1)*100,2)
```

		5	4	3	2	1
Under 40		23.75	41.67	18.77	9.32	6.49
40-49		23.72	39.86	19.13	10.75	6.54
50-59		24.35	39.68	20.14	10.30	5.54
60 or older		25.53	40.41	19.70	9.83	4.52

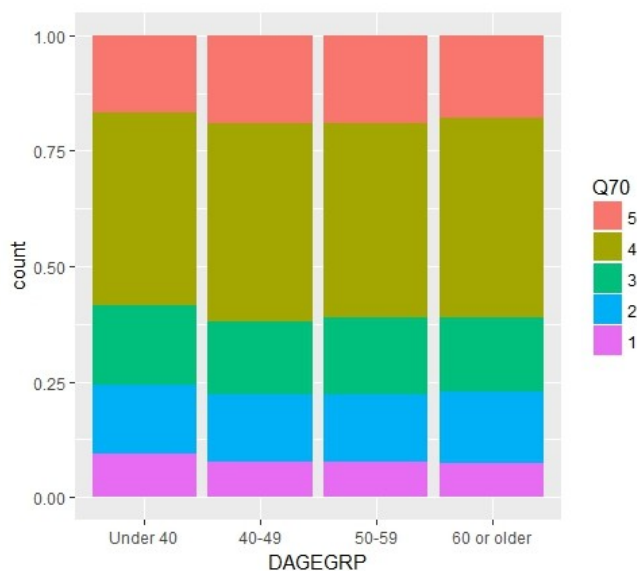
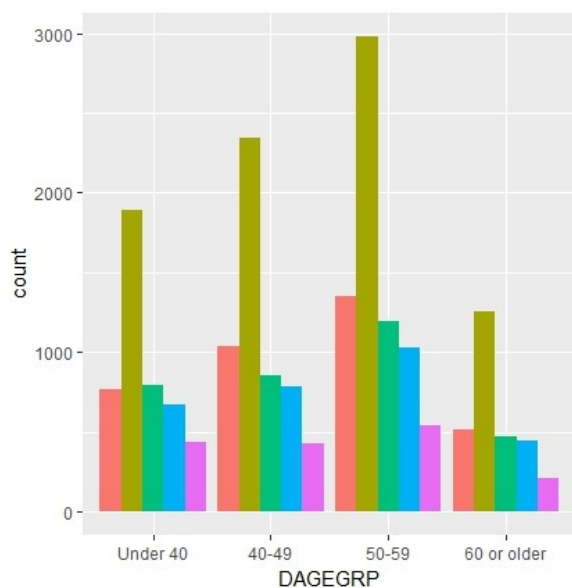
1) Job



```
round(prop.table(tab_age_Q69, 1)*100,2)
```

		5	4	3	2	1
Under 40		19.45	45.46	17.94	11.25	5.90
40-49		21.30	45.63	16.43	10.90	5.75
50-59		23.20	45.10	16.08	10.06	5.56
60 or older		25.85	45.58	15.01	9.73	3.83

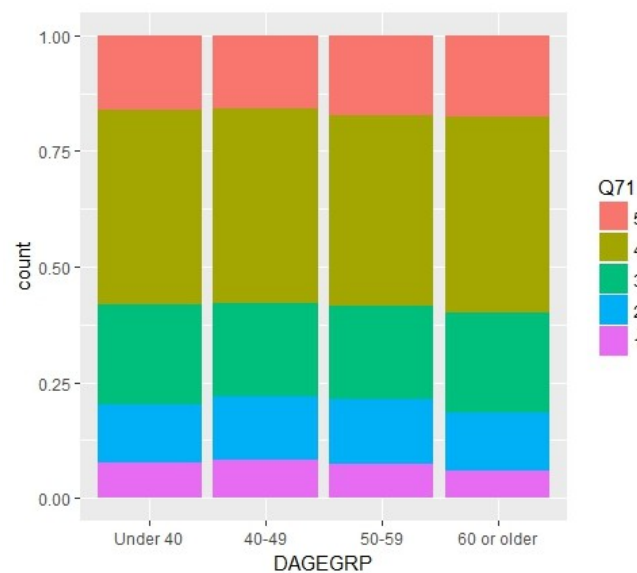
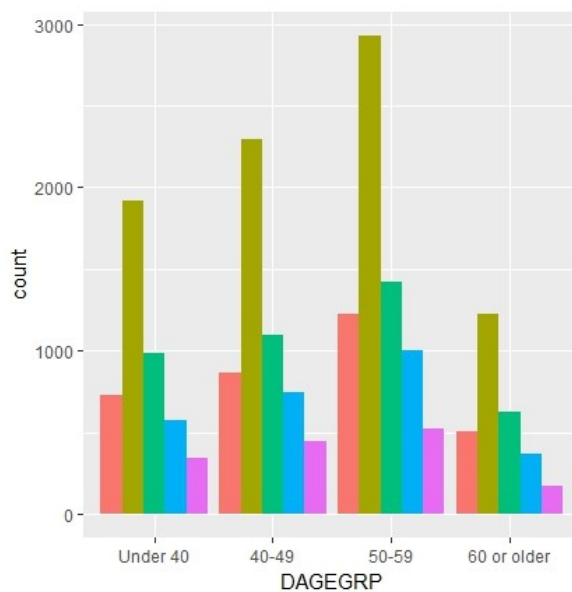
2) Pay



`round(prop.table(tab_age_Q70, 1)*100,2)`

	5	4	3	2	1
Under 40	16.82	41.51	17.37	14.76	9.54
40-49	19.00	43.07	15.64	14.50	7.79
50-59	19.04	41.97	16.80	14.51	7.68
60 or older	17.84	43.27	16.15	15.36	7.38

### 3) Organization



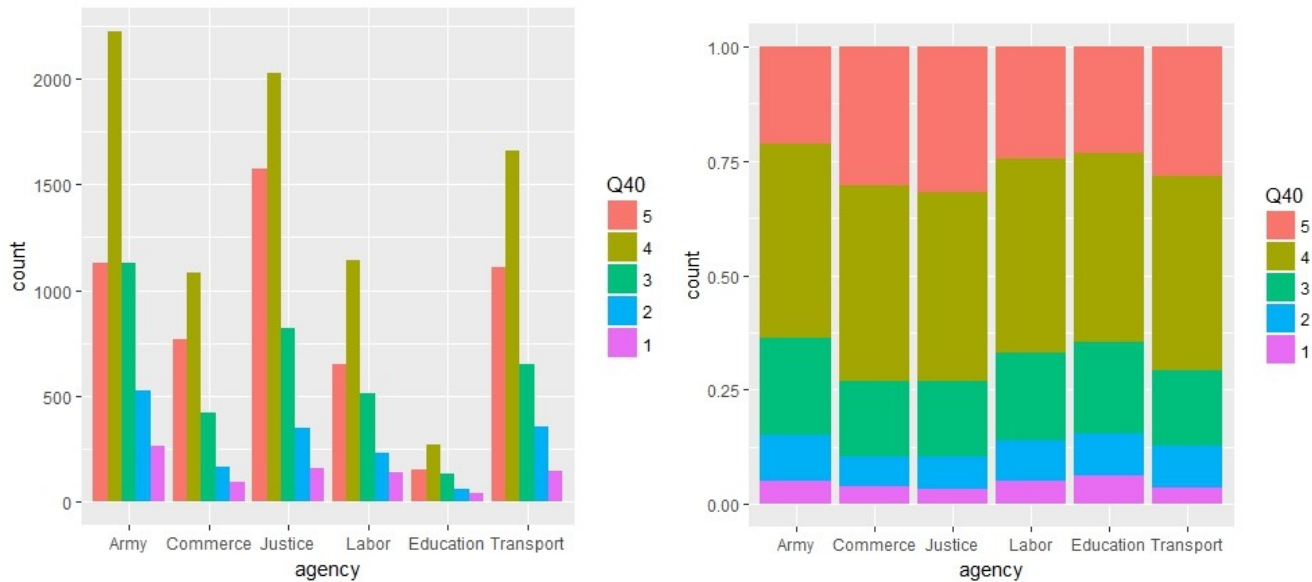
`round(prop.table(tab_age_Q71, 1)*100,2)`

	5	4	3	2	1
Under 40	16.01	42.11	21.62	12.63	7.63
40-49	15.88	42.10	20.16	13.63	8.23
50-59	17.25	41.25	20.00	14.15	7.34
60 or older	17.56	42.27	21.60	12.70	5.87



From the plots, we can see that there are slightly differences among the answers of age groups. Older groups are a little more satisfied with job than younger groups. Under 40 group is less satisfied with pay than other groups.

## Agency

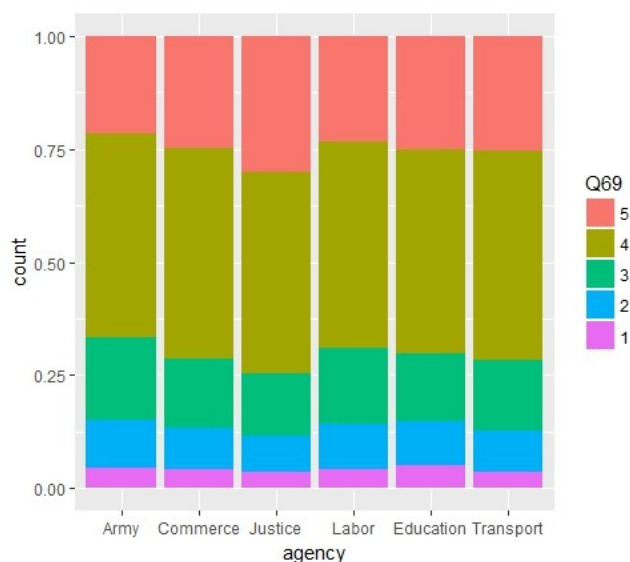
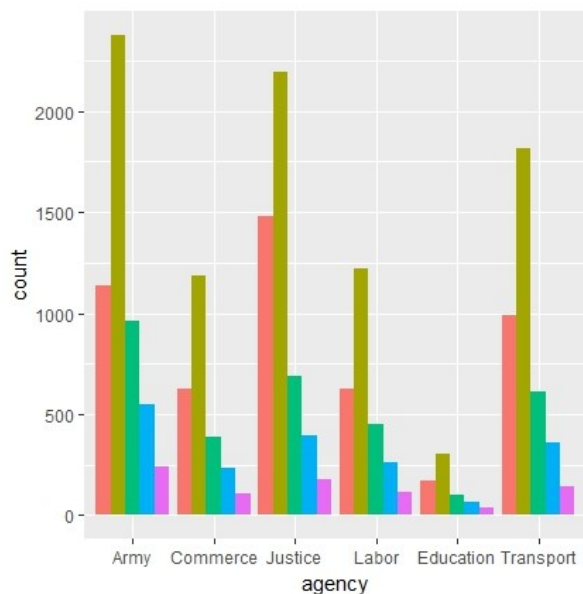


From the plots, we can see that there are distinguish differences among the Q40 answers of agencies. Employees in Departments of Commerce, Justice and Transportation are more satisfied than employees in Departments of Army, Education and Labor. We can get the same result from the table below:

```
round(prop.table(tab_agency, 1)*100,2)
```

	5	4	3	2	1
Army	21.40	42.23	21.38	10.01	4.98
Commerce	30.35	42.72	16.57	6.54	3.82
Justice	31.93	41.05	16.64	7.14	3.24
Labor	24.41	42.58	19.14	8.67	5.20
Education	23.42	41.14	20.12	9.01	6.31
Transport	28.29	42.32	16.56	9.13	3.70

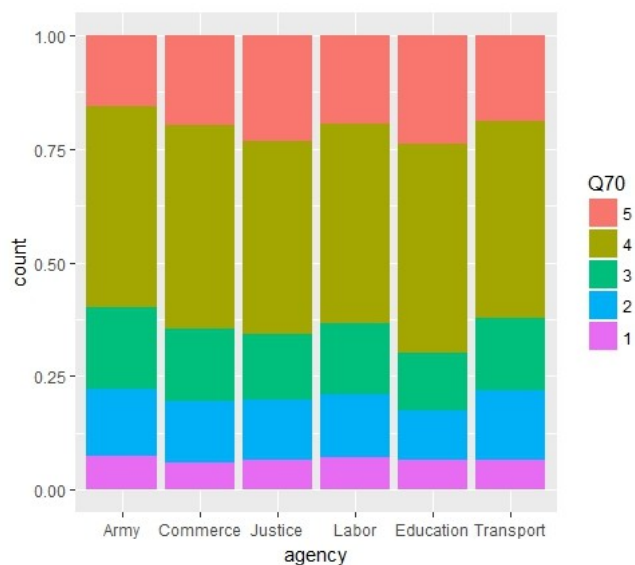
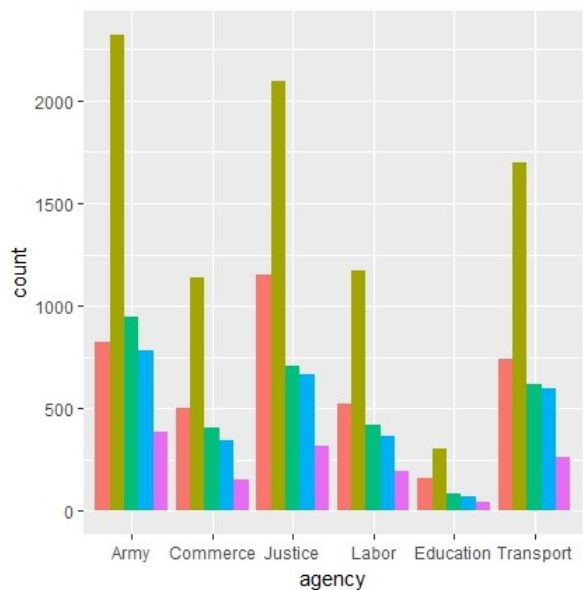
1) Job



`round(prop.table(tab_agency_Q69, 1)*100,2)`

	5	4	3	2	1
Army	21.61	45.16	18.23	10.46	4.54
Commerce	24.69	46.61	15.39	9.21	4.09
Justice	30.04	44.42	13.97	8.01	3.57
Labor	23.33	45.72	16.82	9.79	4.34
Education	25.08	45.20	14.86	9.76	5.11
Transport	25.26	46.25	15.64	9.21	3.65

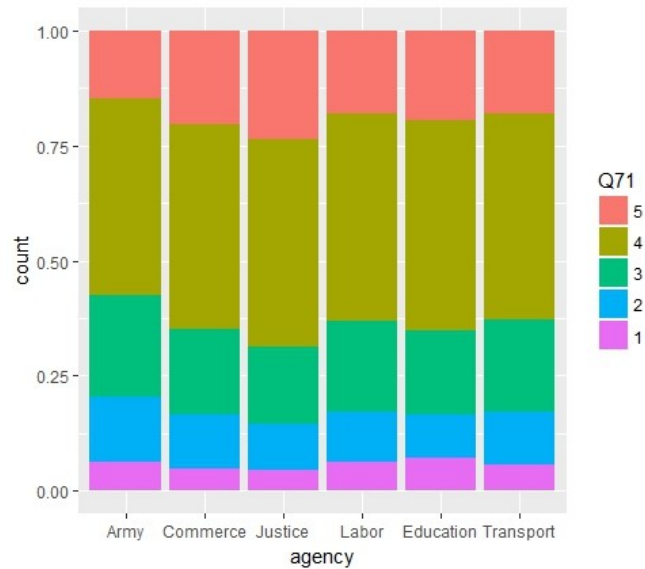
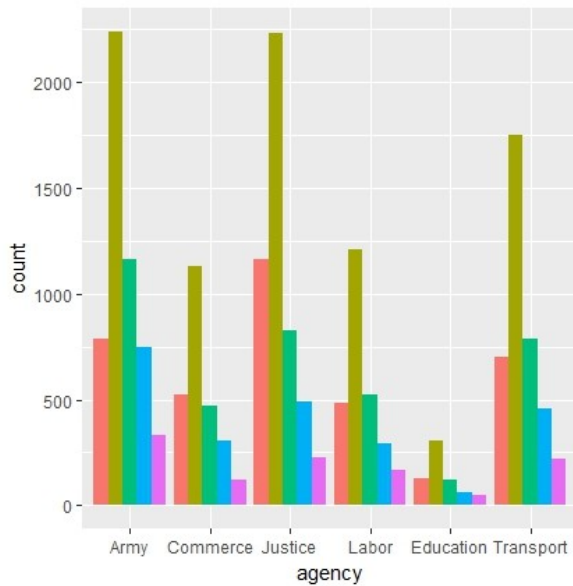
## 2) Pay



```
round(prop.table(tab_agency_Q70, 1)*100,2)
```

	5	4	3	2	1
Army	15.69	44.09	18.02	14.87	7.33
Commerce	19.84	44.76	15.87	13.54	5.98
Justice	23.33	42.47	14.31	13.44	6.45
Labor	19.51	43.85	15.74	13.68	7.21
Education	24.02	45.95	12.61	10.96	6.46
Transport	18.88	43.32	15.84	15.28	6.68

### 3) Organization



```
round(prop.table(tab_agency_Q71, 1)*100,2)
```

	5	4	3	2	1
Army	14.91	42.52	22.07	14.17	6.34
Commerce	20.47	44.37	18.46	12.01	4.69
Justice	23.60	45.17	16.70	9.99	4.54
Labor	18.06	45.16	19.63	10.99	6.17
Education	19.52	45.65	18.17	9.61	7.06
Transport	17.96	44.69	20.10	11.68	5.56

From the plots, we can see that there are distinguish differences among the answers of agencies. Employees in Departments of Justice is the most satisfied with job and organization than employees in other Departments. Department of Army is relatively the least satisfied with job, pay and organization.

## IV. APPROACH

### 1. Data preparation

Data Wrangling

Modeling and Descriptive analysis

Exploratory Data Analysis

Data visualization

Data Story

## **V. DELIVERABLES**

Code, paper, slide and presentation