IBM Attrition Visualization Mini Project

Alan Qin

4/16/2020

Contents

Dataset Introduction

The dataset I am using is the IBM Employee Attrition csv. I downloaded this data set from https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset. This data set is about

- The attrition rate of IBM employees
- The many different factors that may affect attrition rate
- more filler

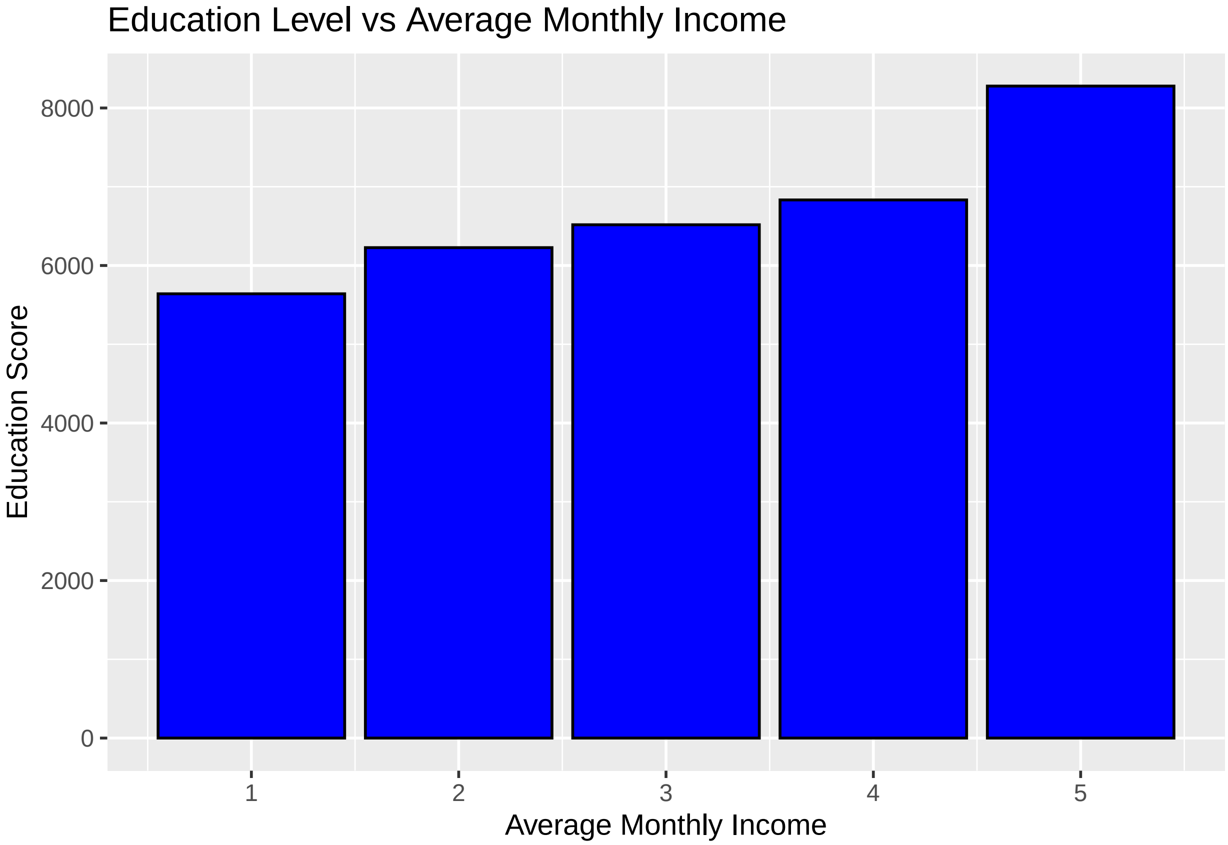
Fixing and Cleaning Data

```r
sum(is.na(ibm.data)) # Make sure no NA's in data
## 0
```

List of Questions and Visualizations

Question 1: How is average income affected by years of Education?

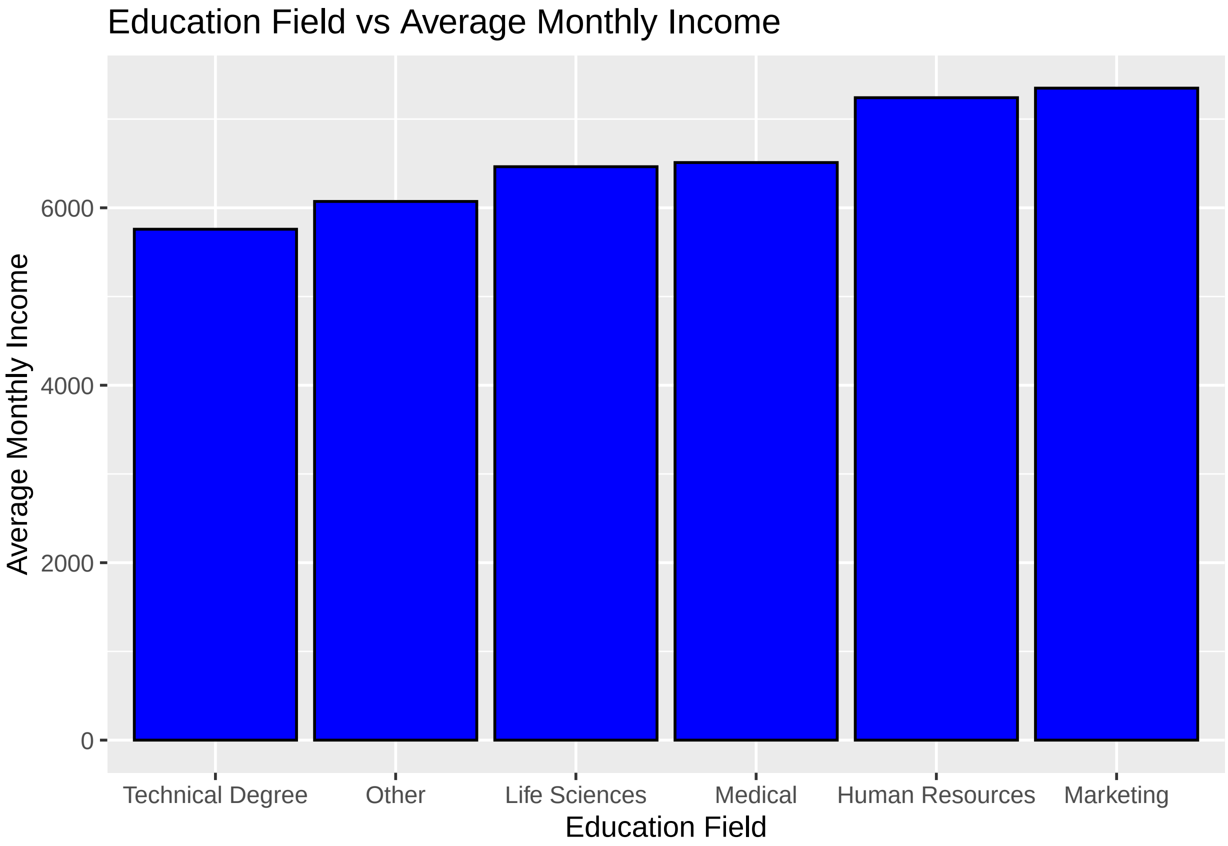
```r
ibm.income = ibm.data %>%
  group_by(Education) %>%
  summarise(average.income = mean(MonthlyIncome))
ggplot(data = ibm.income) +
  geom_bar(aes(x = Education, y = average.income),
           stat = 'identity',
           fill = 'blue',
           color = 'black') +
  labs(x = 'Average Monthly Income', y = 'Education Score') +
  ggtitle('Education Level vs Average Monthly Income')
```



In the IBM Attrition dataset, the education variable is measured with a number 1-5. 1 is below college, 2 is college, 3 is completion of a bachelors degree, 4 is the completion of a masters dgree, and 5 is the completion of a doctorate degree. As we can see from this plot, the average income increases as education increases. One of the most interesting things in this plot is that employees that have an education rating of 1 still make a decent amount of money compared to even a masters degree. Another interesting thing is that doctorate degrees earn much more than a masters degree.

Question 2: How is average income affected by Education Field?

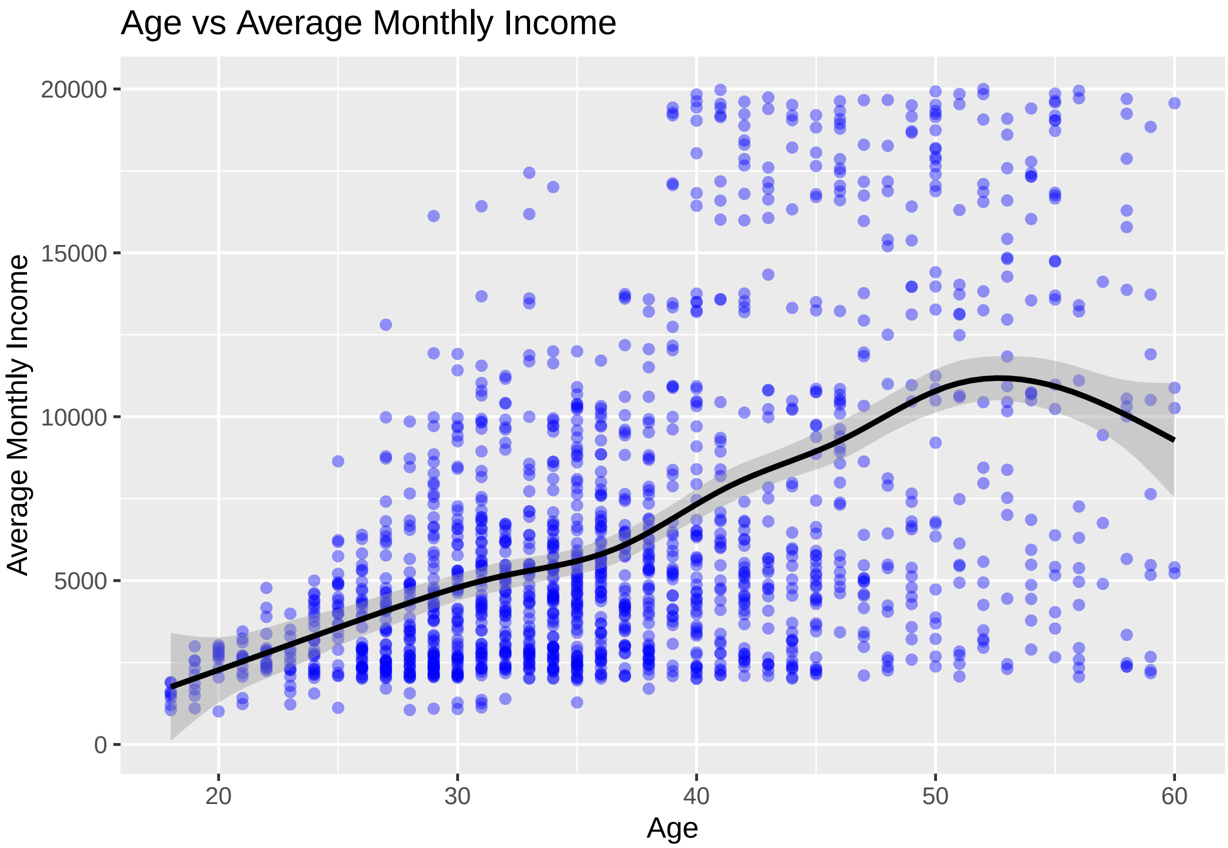
```r
ibm.edu = ibm.data %>% group_by(EducationField) %>% summarise(average.inc = mean(MonthlyIncome))
ggplot(data = ibm.edu) +
  geom_bar(aes(x = reorder(EducationField, average.inc), y = average.inc),
           stat = 'identity',
           fill = 'blue',
           color = 'black') +
  ggtitle('Education Field vs Average Monthly Income') +
  labs(x = 'Education Field', y = 'Average Monthly Income')
```



In the plot above, you can see that there are five general education fields at IBM. The field making the most money is marketing while the field making the least money is a techincal degree. The most interesting thing to me is that medical degrees are in the middle of the pack comparing to other fields. I would say that this data does not make too much sense because I personally think that people in medical fields should be making the highest or one of the highest average incomes.

Question 3: How does age affect income?

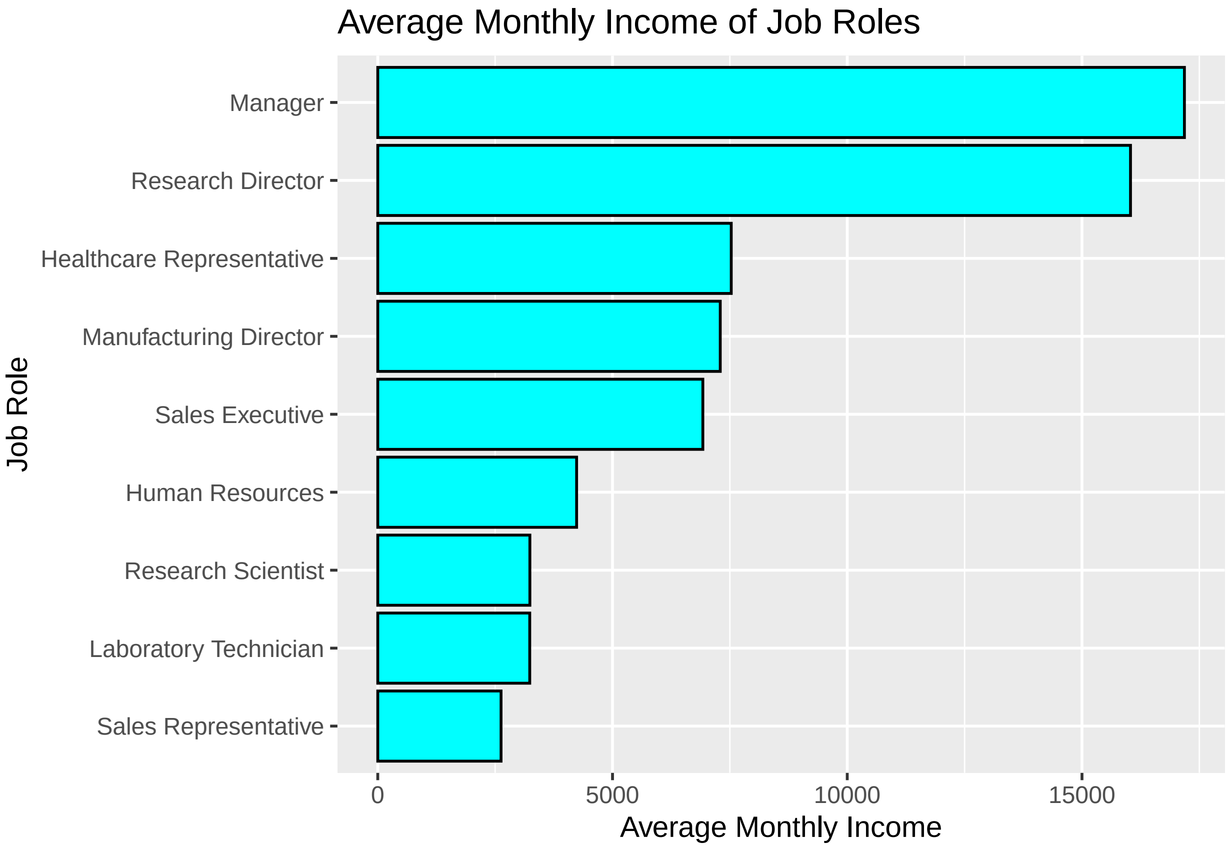
```
ggplot(data = ibm.data) +
  geom_point(aes(x = Age, y = MonthlyIncome), color = 'blue', alpha = .4) +
  ggtitle('Age vs Average Monthly Income') +
  labs(x = "Age", y = 'Average Monthly Income') +
  geom_smooth(aes(x = Age, y = MonthlyIncome), color = 'black')
```



As you can see from this plot, the average monthly income for employees increases as employees get older from ages 18 to around 52. After 52, the average monthly income for employees decreases. I think this situation happens at every company. After around 50, income decreases because employees are not as sharp as they are when they were younger.

Question 4: Job role effect on income?

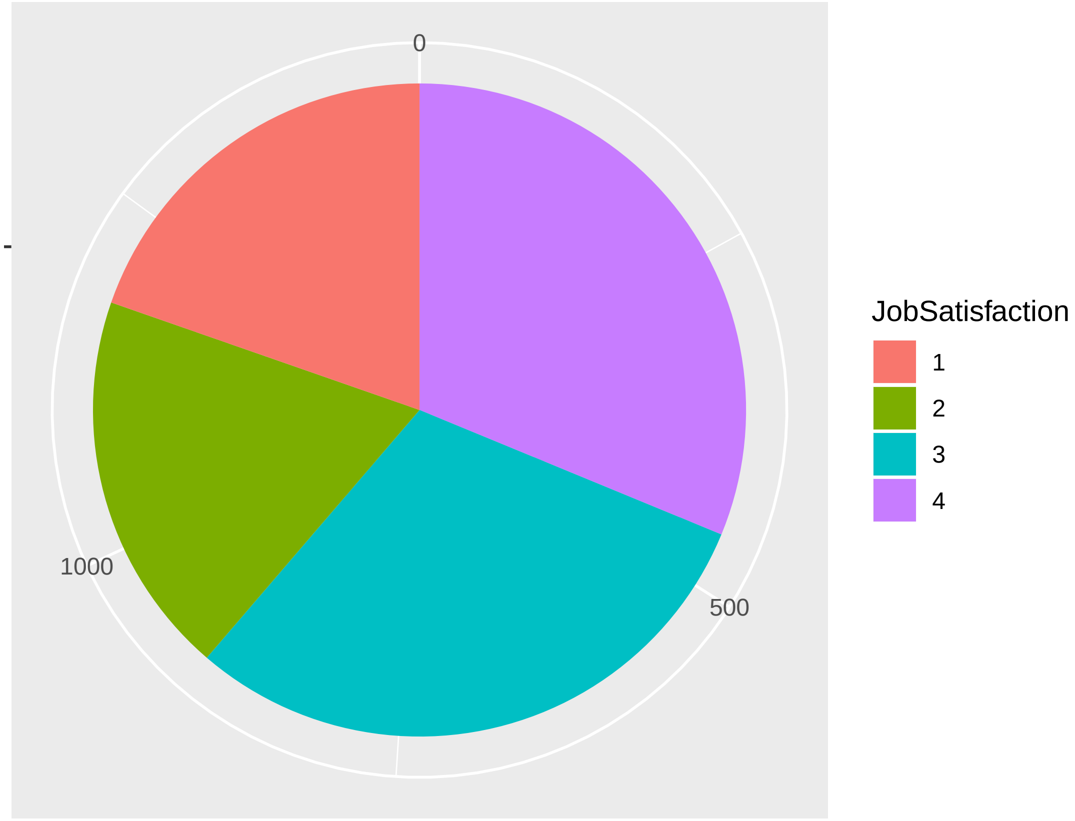
```r
ibm.job = ibm.data %>% group_by(JobRole) %>% summarise(avgInc = mean(MonthlyIncome))
x = ibm.job %>% arrange(desc(avgInc))
ggplot(data = x, aes(x = reorder(JobRole, avgInc), y = avgInc)) +
  geom_bar(stat = 'identity', color = 'black', fill = 'cyan') +
  coord_flip() +
  ggtitle('Average Monthly Income of Job Roles') +
  labs(y = 'Average Monthly Income', x = 'Job Role')
```



In this plot, I believe there weren't too many suprises and the results were relatively predictable. The one thing that suprised me was the fact that Healthcare Representatives made more money than Manufacturing Directors. This is because

Question 5: How happy are employees?

```
pie = ggplot(ibm.data, aes(x = "", fill = as.factor(JobSatisfaction))) +
  geom_bar(width = 1) +
  theme(axis.line = element_blank(), plot.title = element_text(hjust = .5)) +
  labs(fill = 'JobSatisfaction', x=NULL, y = NULL)
pie + coord_polar(theta = 'y', start =0)
```


JobSatisfaction