

# ECO395M STAT LEARNING Homework 1\*

Mingwei Li, Xinyu Leng, Hongjin Long

## Abstract

This document is the first homework of ECO395M STAT LEARNING. The projects **Gas Price, a bike share network, flights at ABIA** and **K-nearest neighbors** are included in this document. The whole project is available at [here](#).

master

1 branch

0 tags

Go to file

Add file

Code

mliw exercise

8ba4b15 2 hours ago 4 commits

exercise	exercise	2 hours ago
Rmd的基本介绍.txt	exercise	2 hours ago
hk1.Rmd	1	yesterday
hk1.pdf	1	yesterday

Help people interested in this repository understand your project by adding a README.

Add a README

---

\*Mingwei Li, Xinyu Leng and Hongjin Long are master students of economics, The University of Texas at Austin

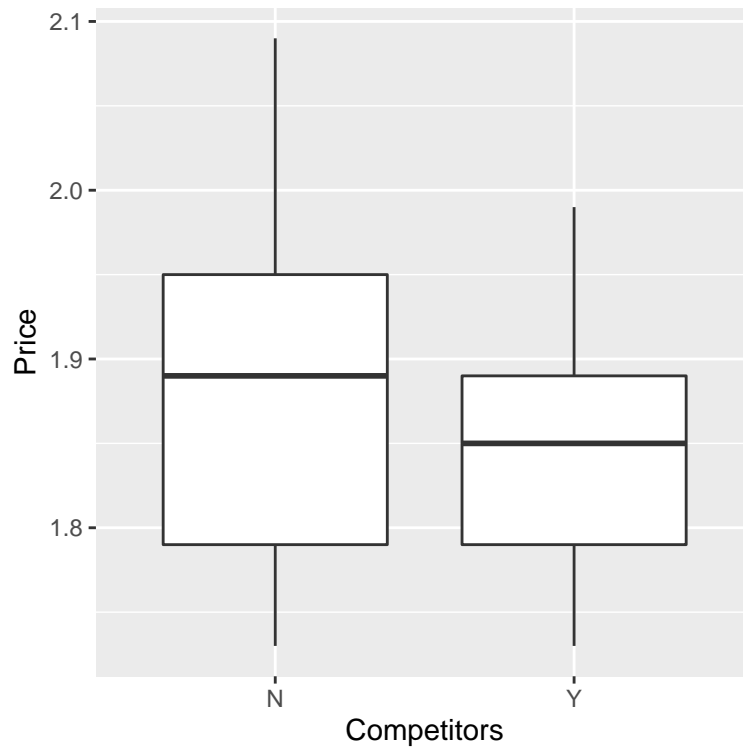
# Contents

<b>1</b>	<b>Gas Price</b>	<b>3</b>
<b>2</b>	<b>a bike share network</b>	<b>8</b>

# 1 Gas Price

(A) Gas stations charge more if they lack direct competition in sight (boxplot).

```
p0 = ggplot(data=gasprice) +  
  geom_boxplot(aes(x=Competitors, y=Price))  
p0
```

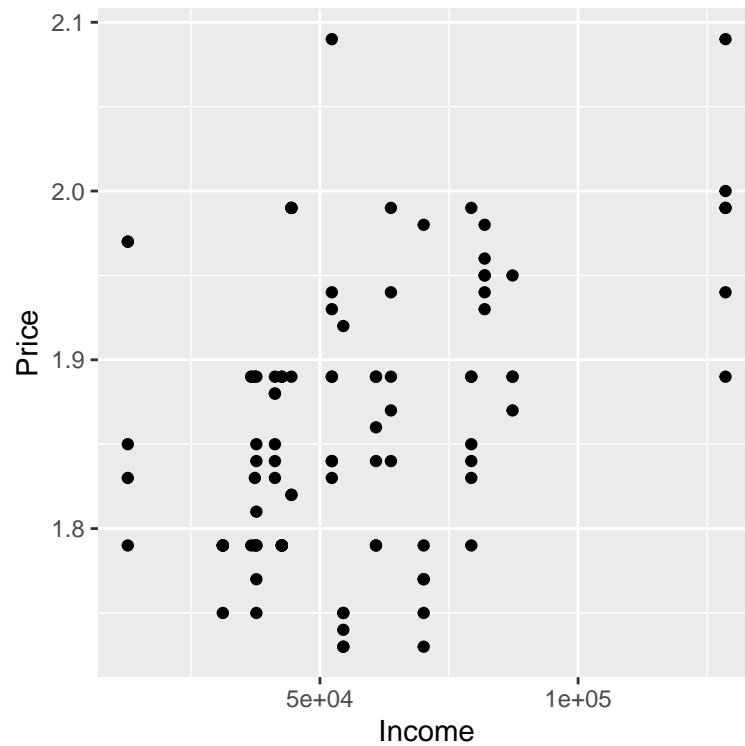


**Claim:** Gas stations charge more if they lack direct competition in sight

**Conclusion:** The theory is supported by the data. As the left box is higher than the right box.

(B) The richer the area, the higher the gas price (scatter plot).

```
p0 = ggplot(data=gasprice) +  
  geom_point(aes(x=Income, y=Price))  
p0
```

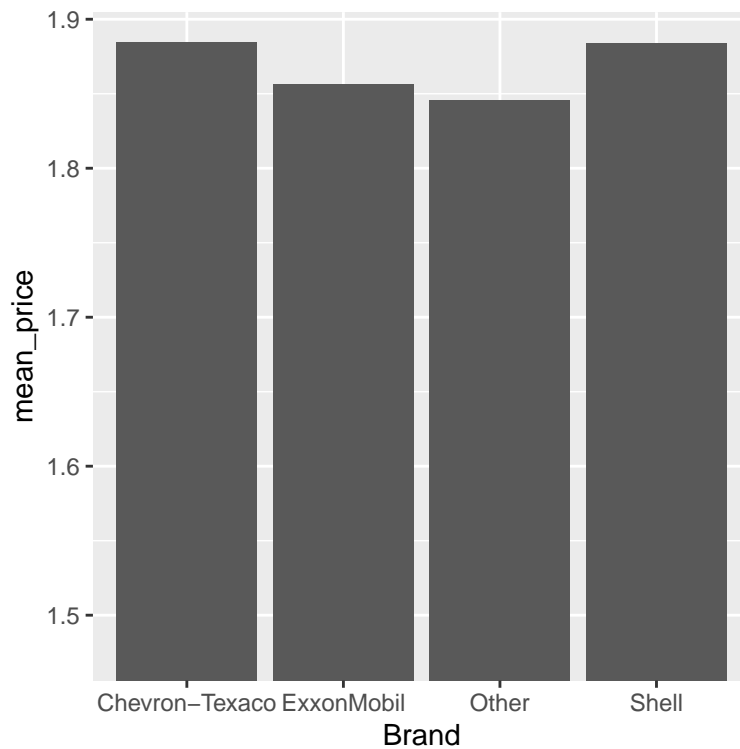


**Claim:** The richer the area, the higher the gas price

**Conclusion:** The theory is supported by the data. As the points in the figure show an increasing trend (this trend not very clear).

(C) Shell charges more than other brands (bar plot).

```
d4 = gasprice %>%  
  group_by(Brand) %>%  
  summarize(mean_price=mean(Price))  
p0 = ggplot(data=d4) +  
  geom_col(aes(x=Brand, y=mean_price))  
p0 + coord_cartesian(ylim =c(min(d4$mean_price)*0.8, max(d4$mean_price)))
```

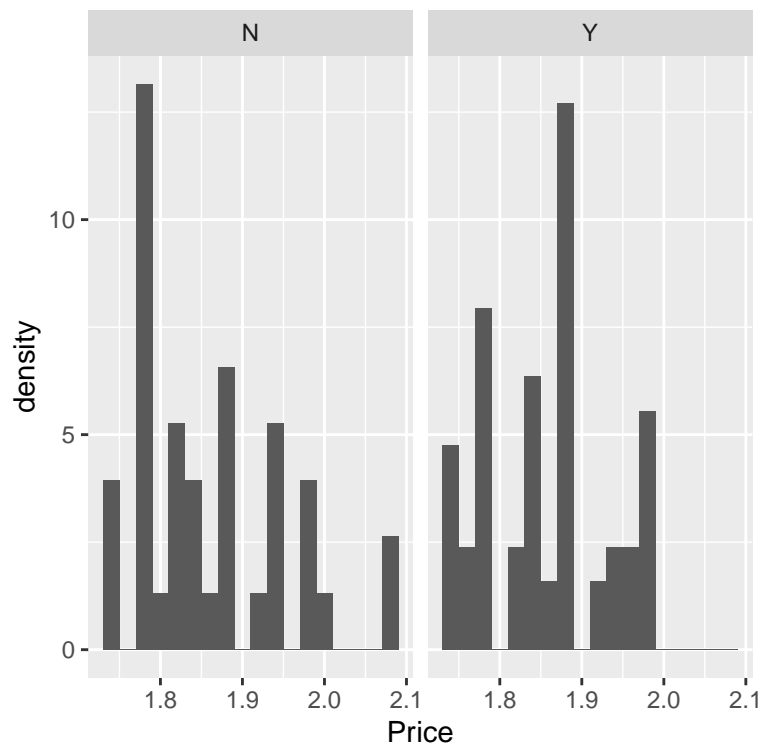


**Claim:** Shell charges more than other brands

**Conclusion:** The theory is supported by the data. As the bar of Shell is the highest (the same as ChevronTexaco).

(D) Gas stations at stoplights charge more (faceted histogram).

```
p0 = ggplot(data=gasprice) +  
  geom_histogram(aes(x=Price, after_stat(density)), binwidth=0.02) +  
  facet_wrap(~Stoplight)  
p0
```



**Claim:** Gas stations at stoplights charge more

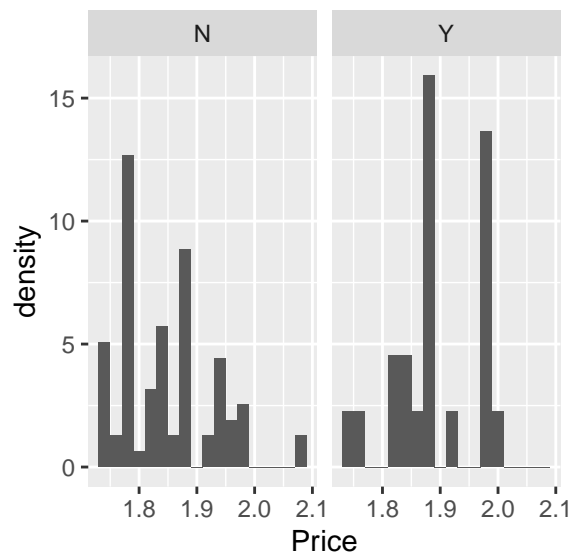
**Conclusion:** The theory is supported by the data. As more density is put on the right tail of gas stations at stoplights(However, the graph is not very clear in this problem)

(E) Gas stations with direct highway access charge more (your choice of plot).

faceted histogram and boxplot are used to solve this problem.

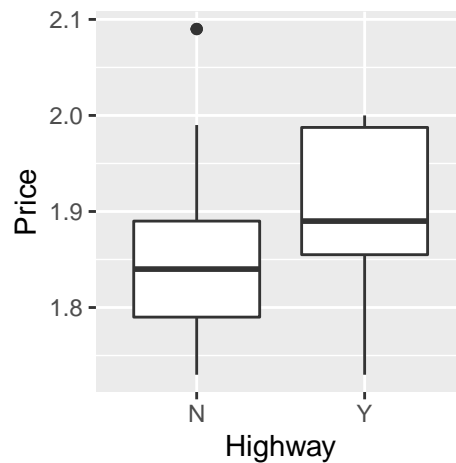
(1) faceted histogram:

```
p0 = ggplot(data=gasprice) +  
  geom_histogram(aes(x=Price, after_stat(density)), binwidth=0.02) +  
  facet_wrap(~Highway)  
p0
```



(2) boxplot:

```
p0 = ggplot(data=gasprice) +  
  geom_boxplot(aes(x=Highway, y=Price))  
p0
```



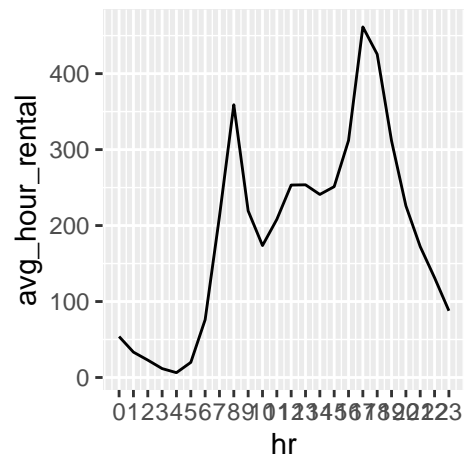
**Claim:** Gas stations with direct highway access charge more

**Conclusion:** The theory is supported by the data. As in histogram, more density is put on the right tail of gas stations with direct highway access. The boxplot reaches the same conclusion.

## 2 a bike share network

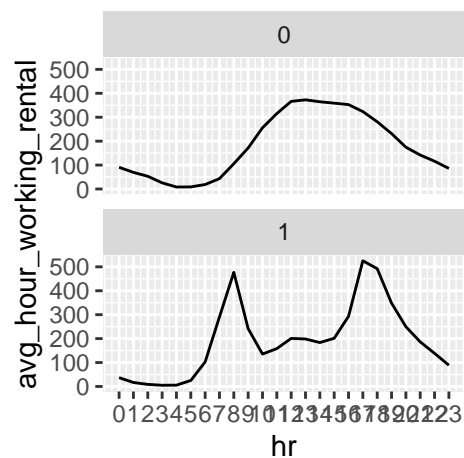
(a) Plot A: a line graph showing average bike rentals (total) versus hour of the day (hr).

```
avg_hr_rent=bikeshare %>%  
  group_by(hr) %>%  
  summarize(avg_hour_rental=mean(total))  
ggplot(data=avg_hr_rent) +  
  geom_line(aes(x=hr, y=avg_hour_rental)) +  
  scale_x_continuous(breaks=0:23)
```



(b) Plot B: a faceted line graph showing average bike rentals versus hour of the day, faceted according to whether it is a working day (workingday).

```
avg_hr_work_rent=bikeshare %>%  
  group_by(hr, workingday) %>%  
  summarize(avg_hour_working_rental=mean(total))  
ggplot(data=avg_hr_work_rent) +  
  geom_line(aes(x=hr, y=avg_hour_working_rental)) +  
  scale_x_continuous(breaks=0:23) +  
  facet_wrap(~ workingday, nrow=2)
```





(c) Plot C: a faceted bar plot showing average ridership during the 8 AM hour by weather situation code (weathersit), faceted according to whether it is a working day or not.

```
bikeshare_8am=bikeshare %>%  
  filter(hr==8)  
bike_8am_work_wather= bikeshare_8am %>%  
  group_by(workingday, weathersit) %>%  
  summarise(avg_8am_working_wather=mean(total))  
ggplot(data=bike_8am_work_wather) +  
  geom_col(aes(x=weathersit, y=avg_8am_working_wather)) +  
  facet_wrap(~ workingday, nrow=2)
```

