

Statistics 135 – Lab Project

April 26, 2015

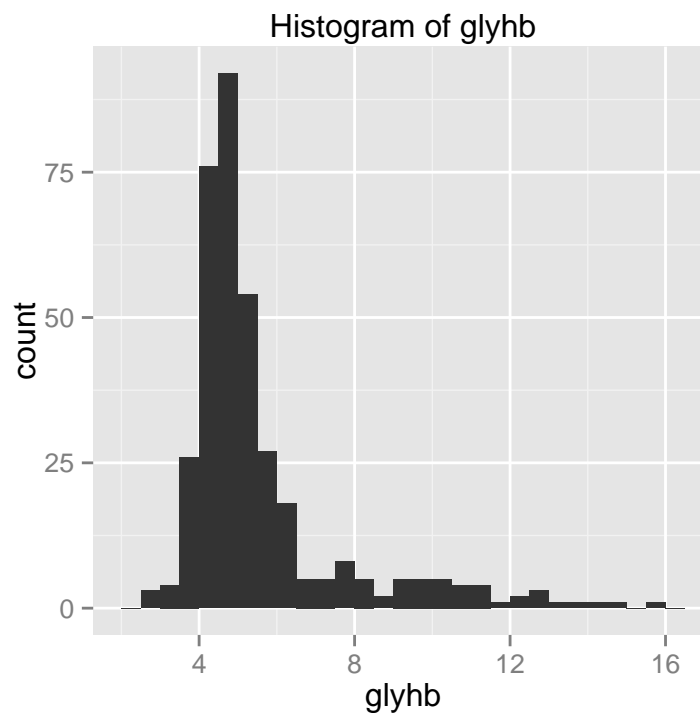
1 Background

```
library("ggplot2")
library("grid")
library("gridExtra")

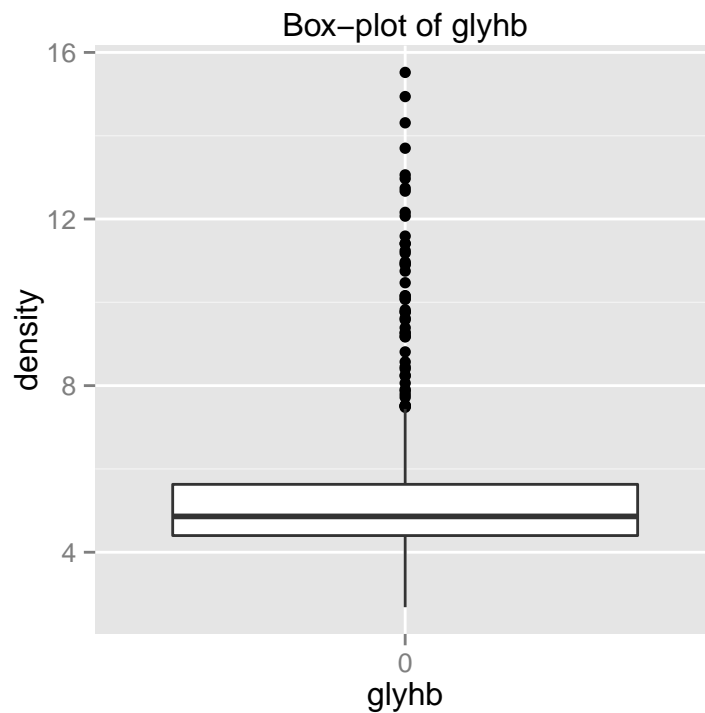
data.df <- na.omit(read.csv("diabetes.csv"))
```

2 Accessing Data, Visualization and Summarization

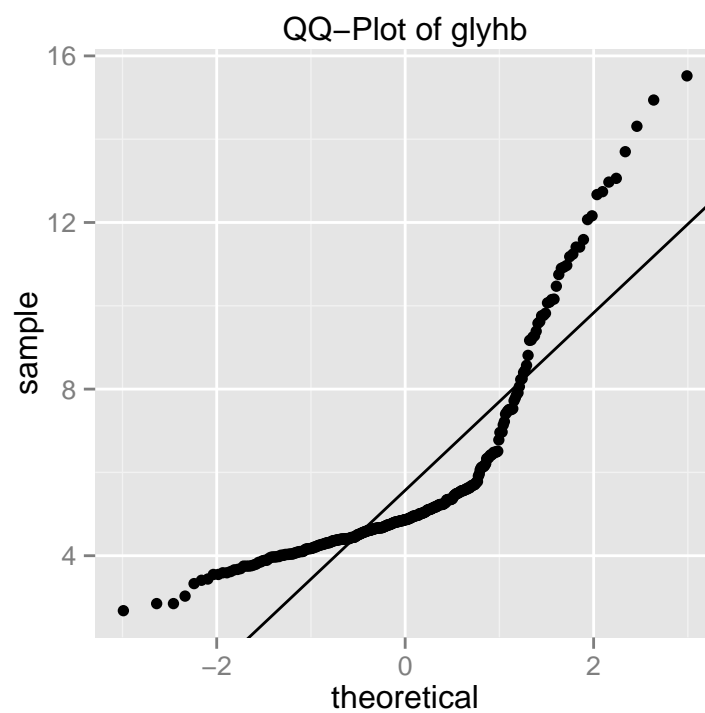
```
1. ggplot(data.df) +
  geom_histogram(aes(x=glyhb), binwidth=0.5) +
  labs(title="Histogram of glyhb") +
  theme(plot.title=element_text(size=rel(1)))
```



```
ggplot(data.df) +
  geom_boxplot(aes(x=factor(0), y=glyhb)) +
  labs(title="Box-plot of glyhb", x="glyhb", y="density") +
  theme(plot.title=element_text(size=rel(1)))
```

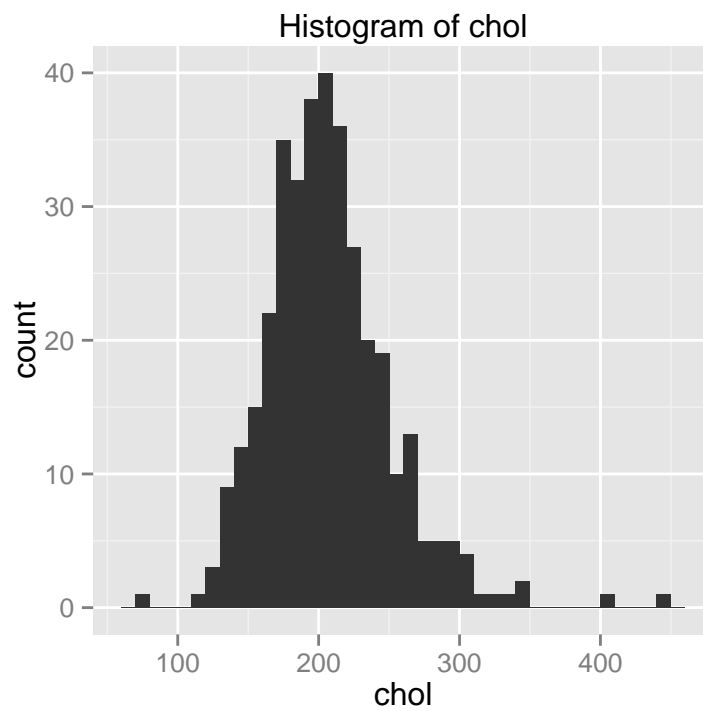


```
ggplot(data.df) +
  stat_qq(aes(sample=glyhb)) +
  geom_abline(aes(intercept=mean(glyhb), slope=sd(glyhb))) +
  labs(title="QQ-Plot of glyhb") +
  theme(plot.title=element_text(size=rel(1)))
```

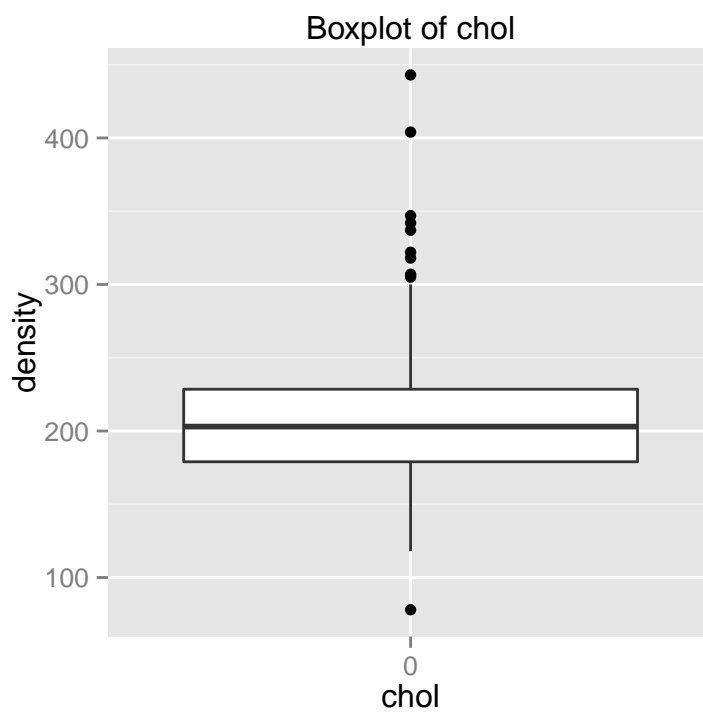


```
2. ggplot(data.df) +
  geom_histogram(aes(x=chol), binwidth=10) +
  labs(title="Histogram of chol") +
```

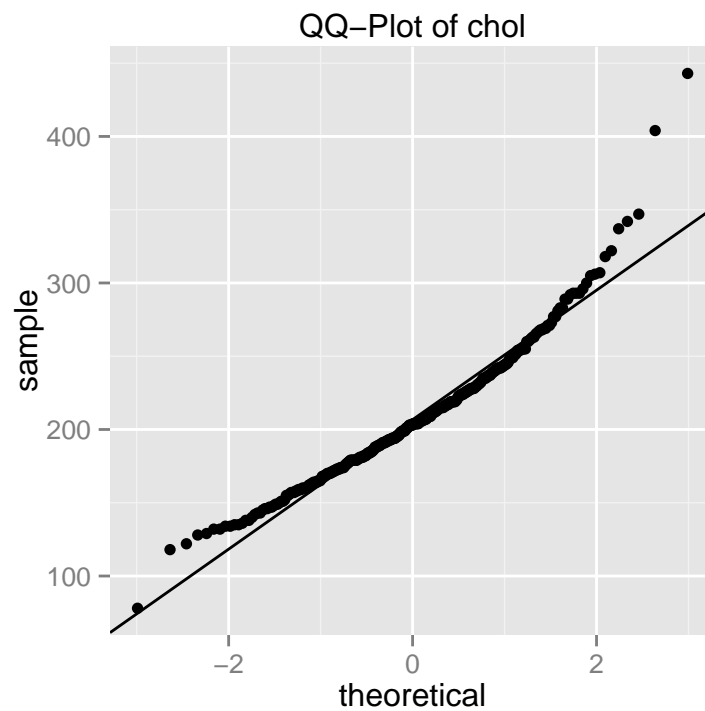
```
theme(plot.title=element_text(size=rel(1)))
```



```
ggplot(data.df) +
  geom_boxplot(aes(x=factor(0), y=chol)) +
  labs(title="Boxplot of chol", x="chol", y="density") +
  theme(plot.title=element_text(size=rel(1)))
```

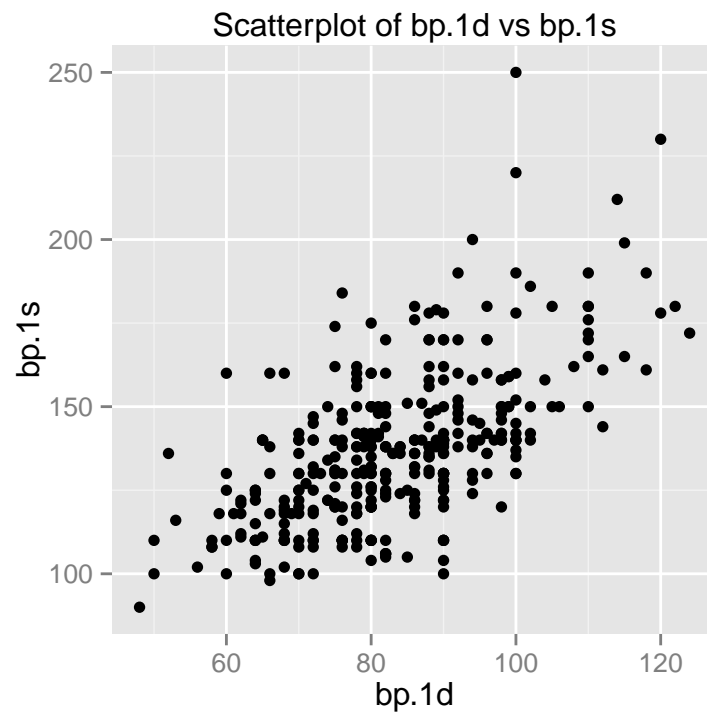


```
ggplot(data.df) +
  stat_qq(aes(sample=chol)) +
  geom_abline(aes(intercept=mean(chol), slope=sd(chol))) +
  labs(title="QQ-Plot of chol") +
  theme(plot.title=element_text(size=rel(1)))
```

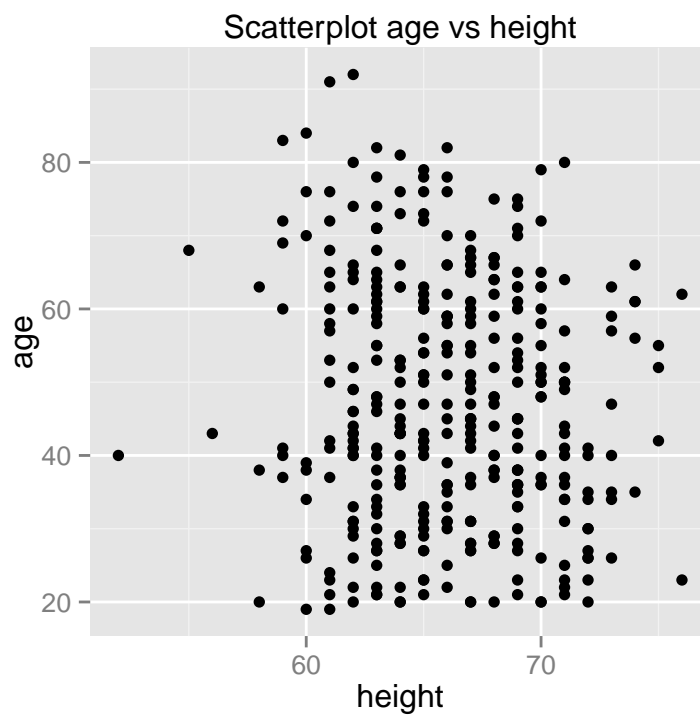


3.

```
ggplot(data.df) +
  geom_point(aes(x=bp.1d, y=bp.1s)) +
  labs(title="Scatterplot of bp.1d vs bp.1s") +
  theme(plot.title=element_text(size=rel(1)))
```



```
ggplot(data.df) +  
  geom_point(aes(x=height, y=age)) +  
  labs(title="Scatterplot age vs height") +  
  theme(plot.title=element_text(size=rel(1)))
```



- 4.
- 5.
- 6.

3 Parametric Inferece

- 1.
- 2.
- 3.

4 Testing

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.

5 Regression

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.