

Introduction to EDA

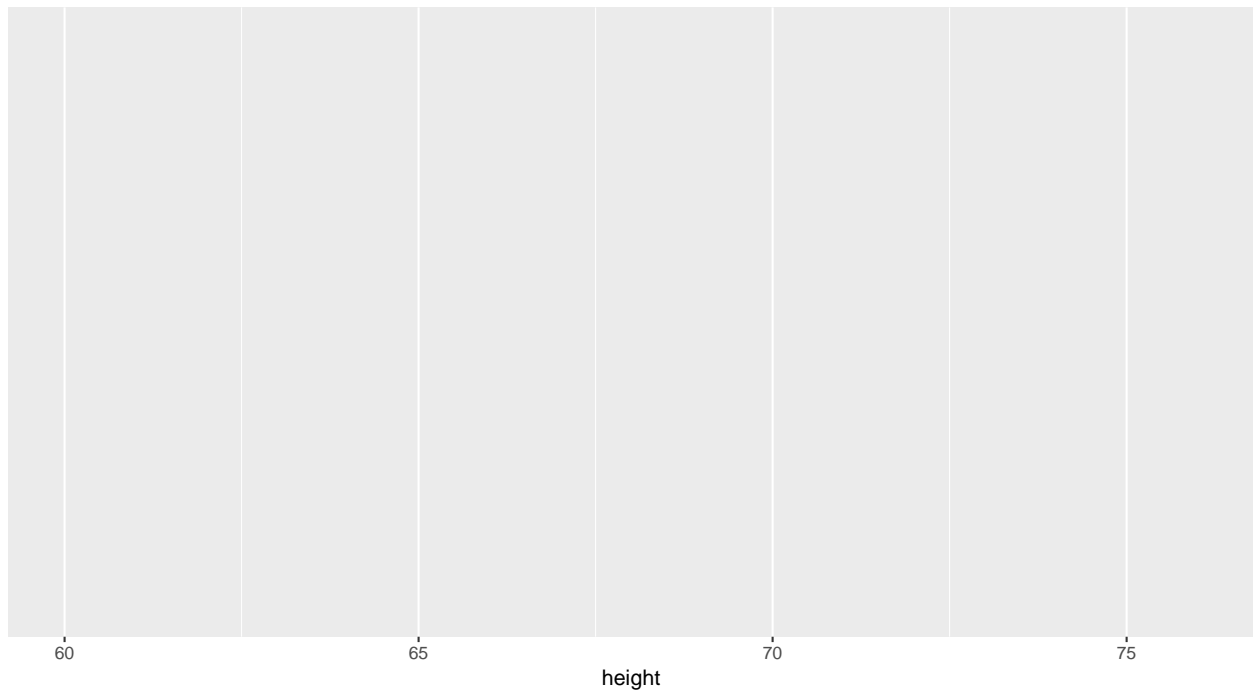
Pramod Duvvuri

3/11/2019

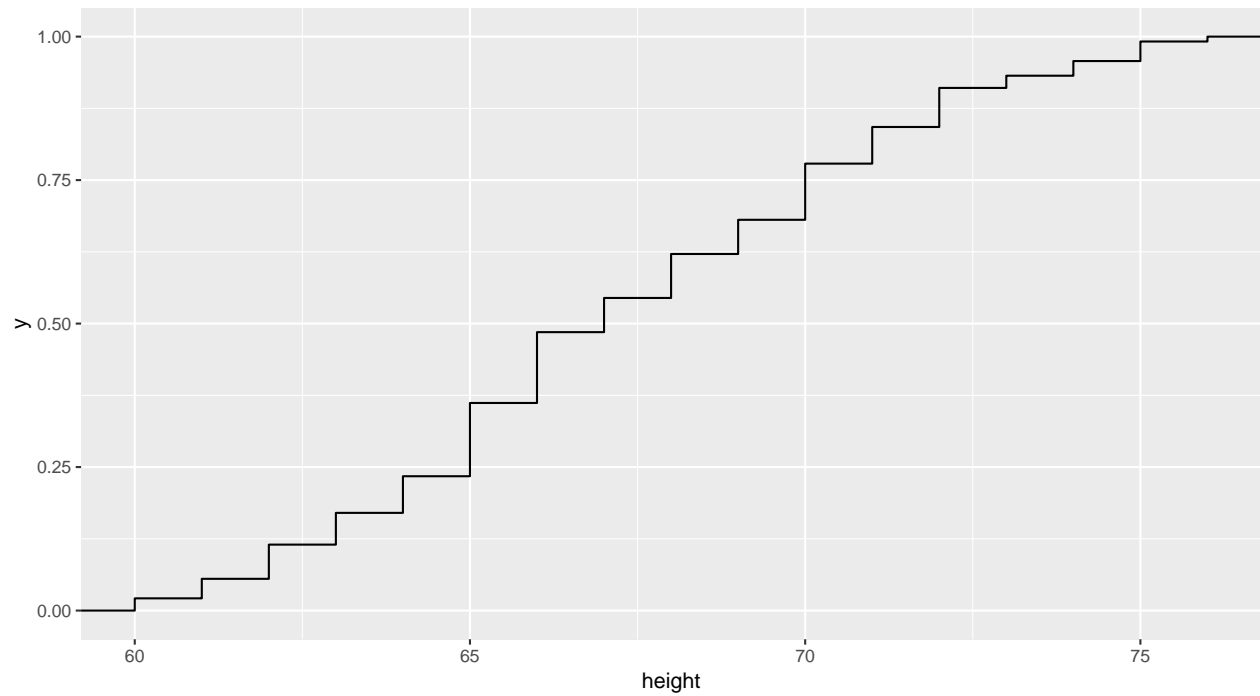
Learning ggplot2

```
# Load installed packages  
library('lattice')  
library('ggplot2')
```

```
ggplot(singer, aes(x = height))
```

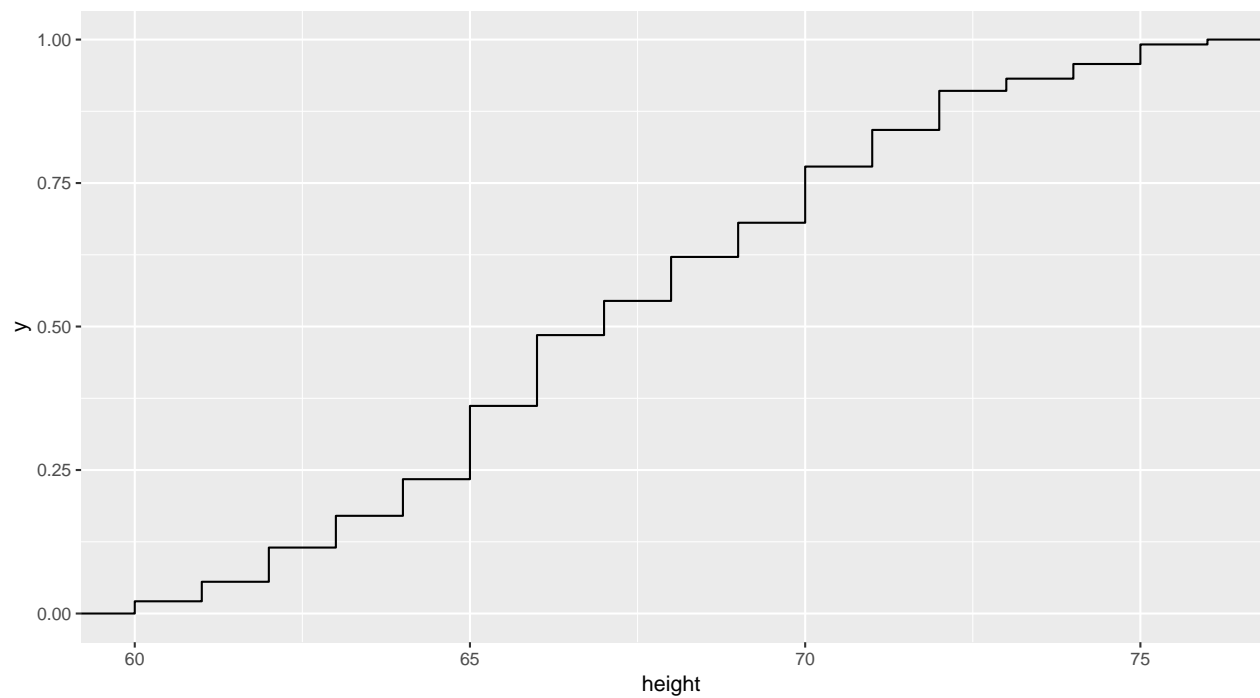


```
# ECDF in ggplot2  
ggplot(singer, aes(x = height)) + stat_ecdf()
```

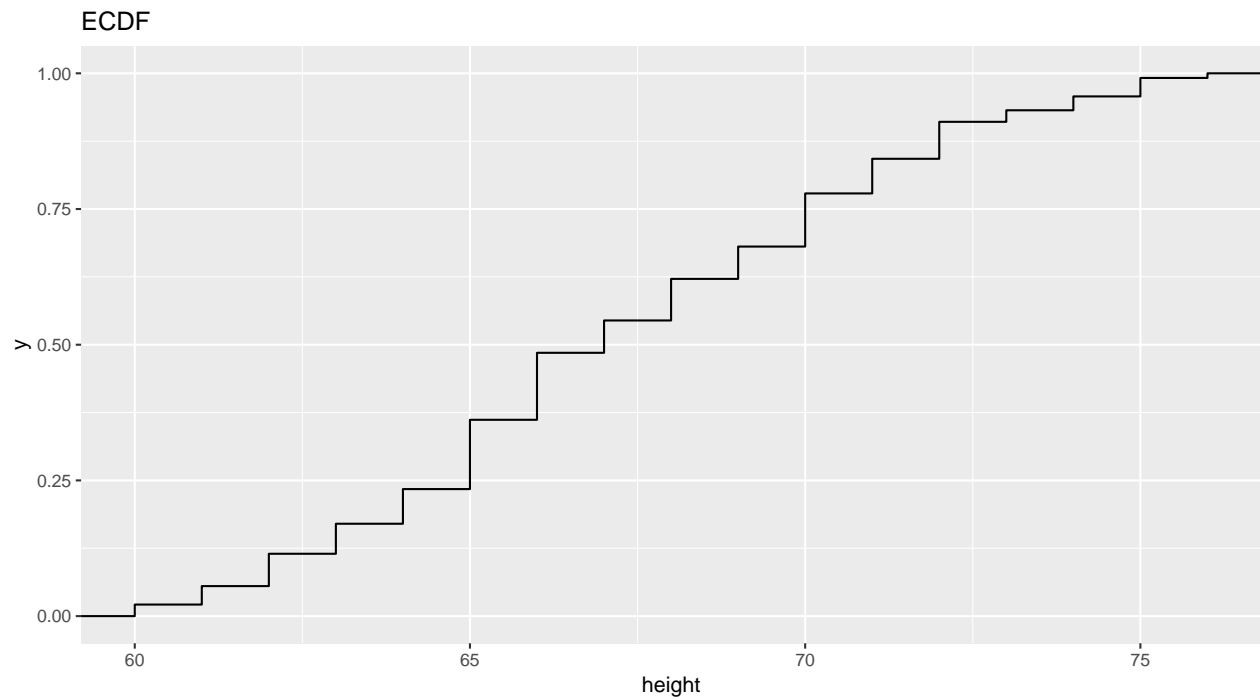


Basic Plots

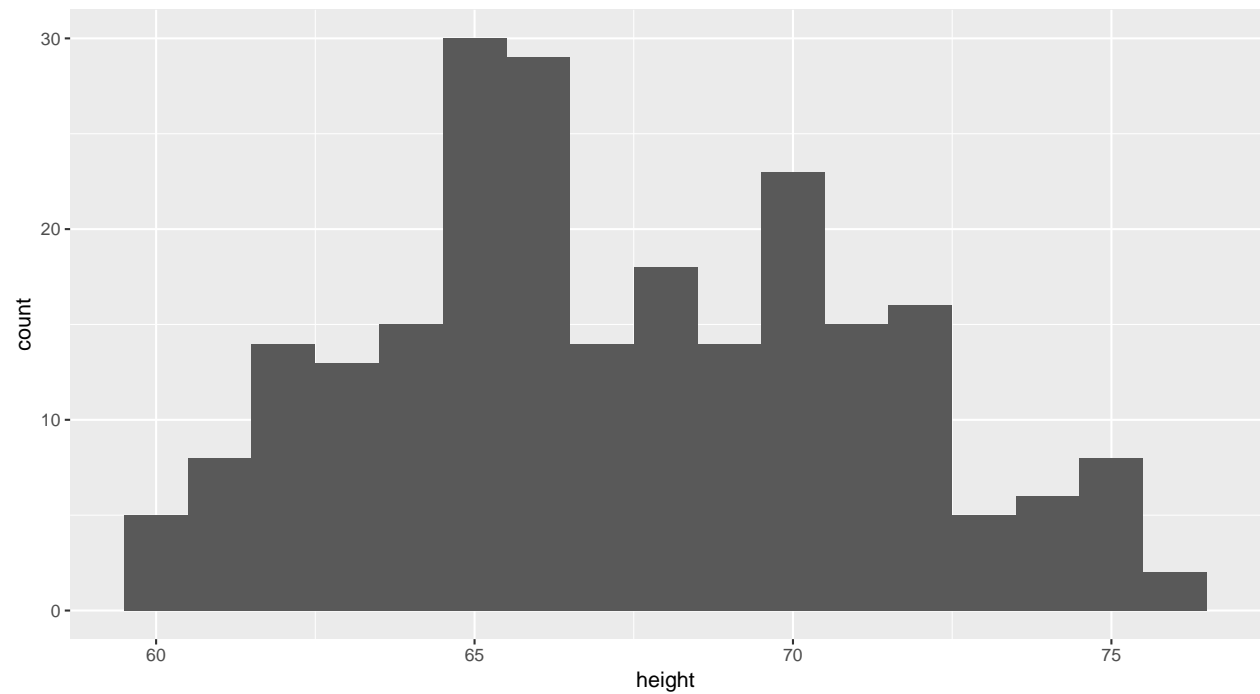
```
# Alternate way to get ECDF  
singer_gg <- ggplot(singer, aes(x = height))  
singer_gg + stat_ecdf()
```



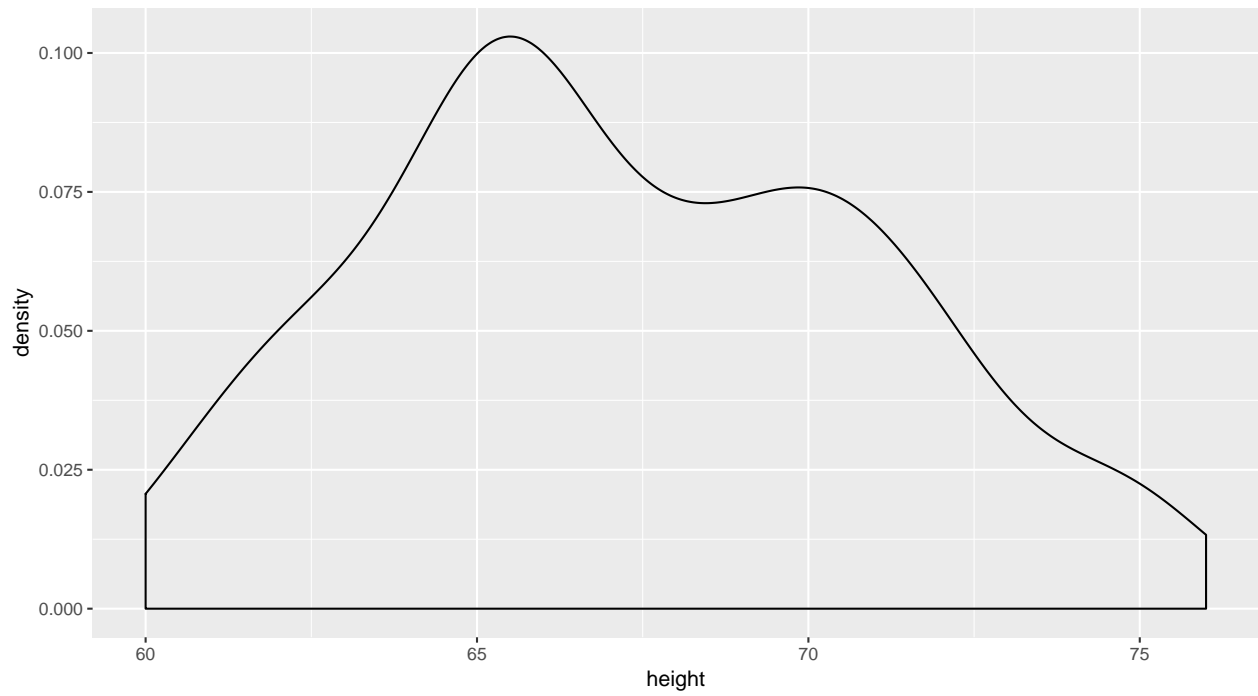
```
# Add Labels  
singer_gg + stat_ecdf() + ggtitle('ECDF')
```



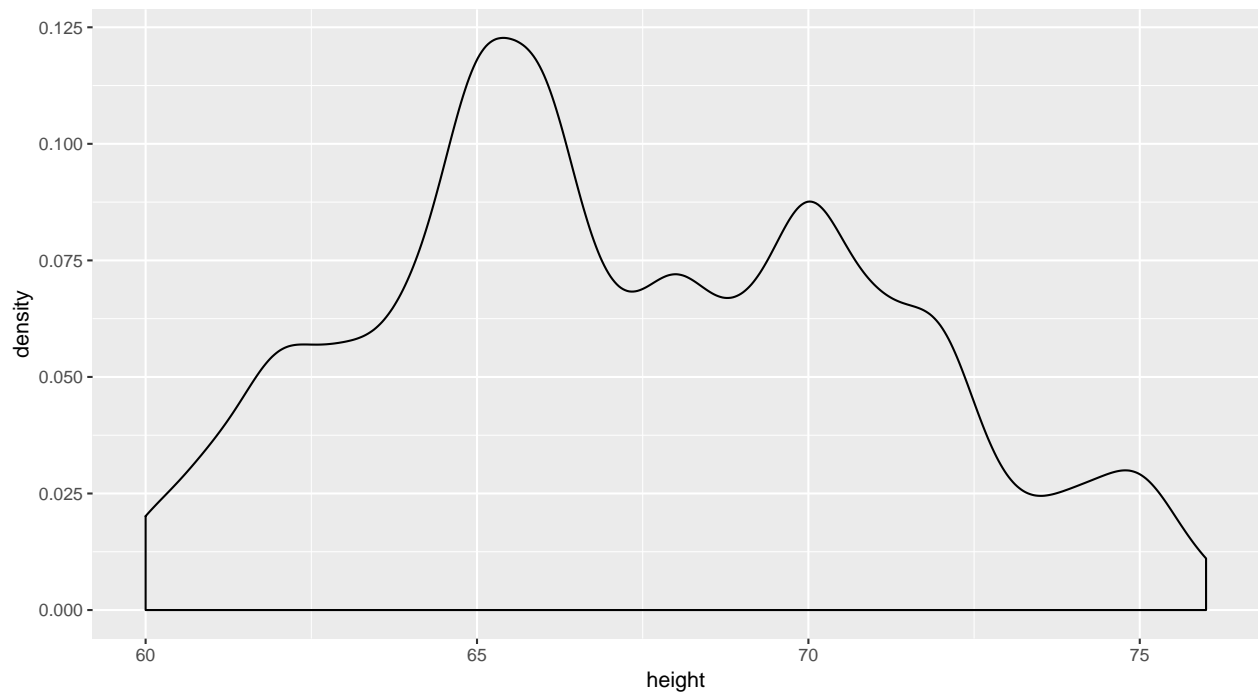
```
# Histogram  
#ggplot(singer, aes(x= height)) + geom_histogram()  
ggplot(singer, aes(x= height)) + geom_histogram(binwidth = 1)
```



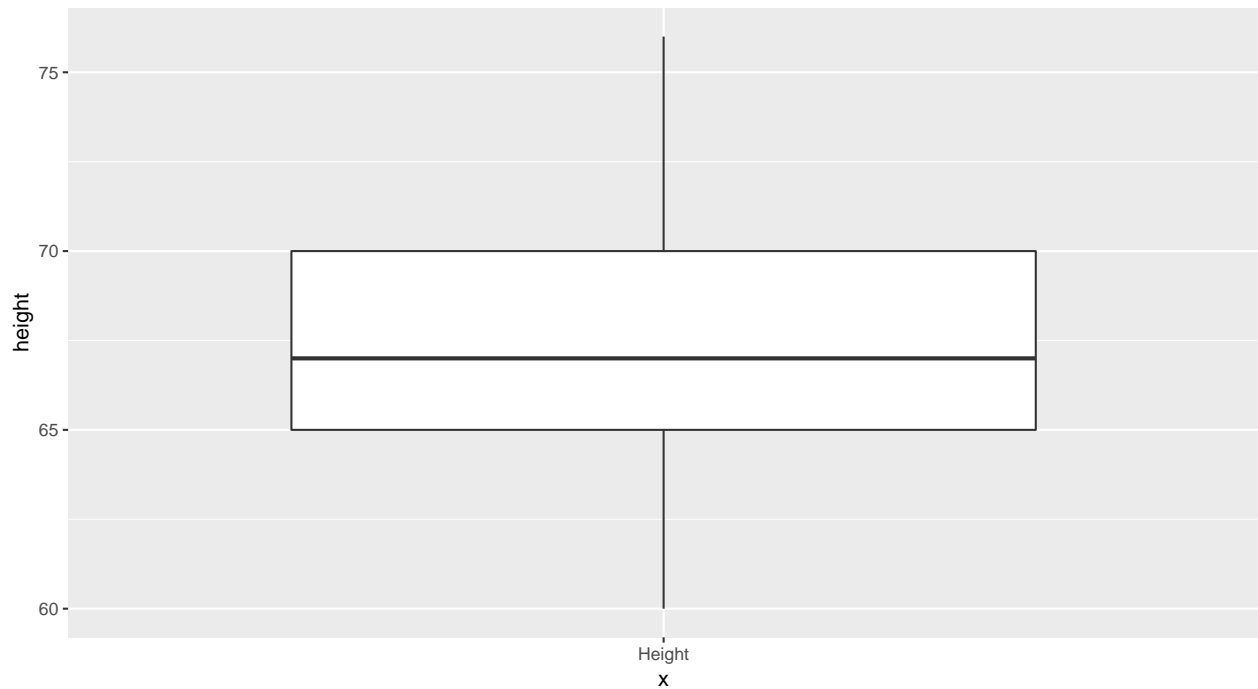
```
# Density Plot  
ggplot(singer, aes(x= height)) +  
  geom_density()
```



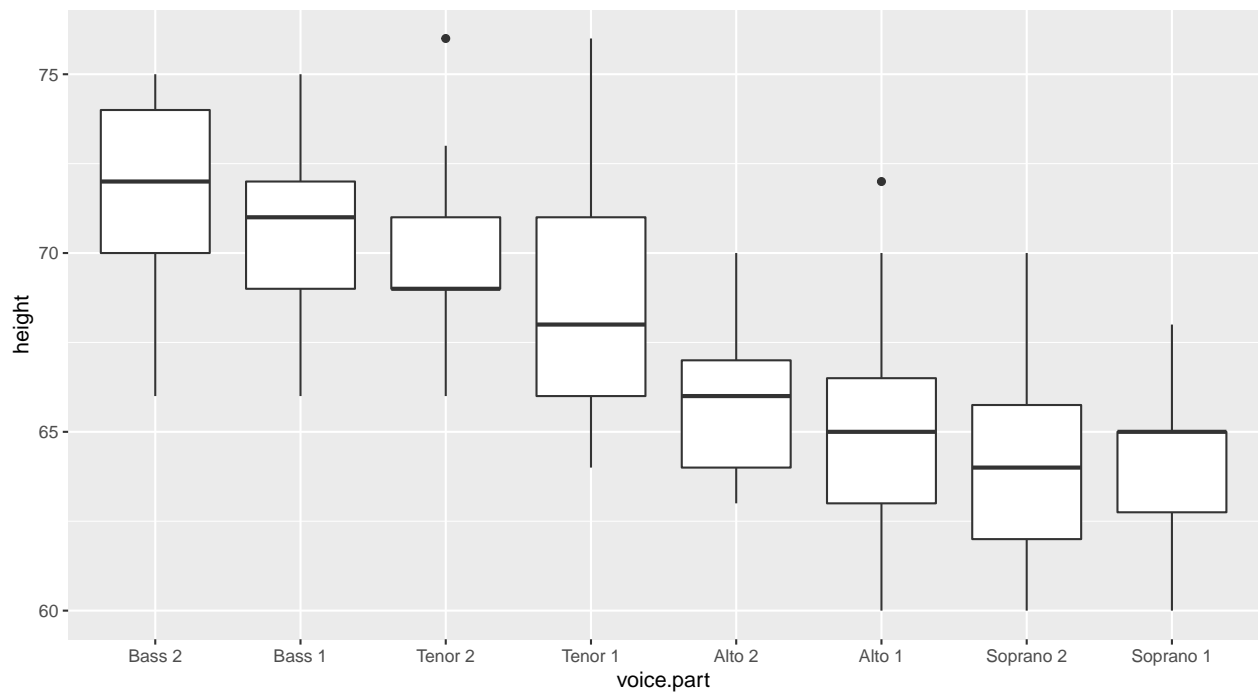
```
ggplot(singer, aes(x= height)) +  
  geom_density(adjust = 0.5)
```



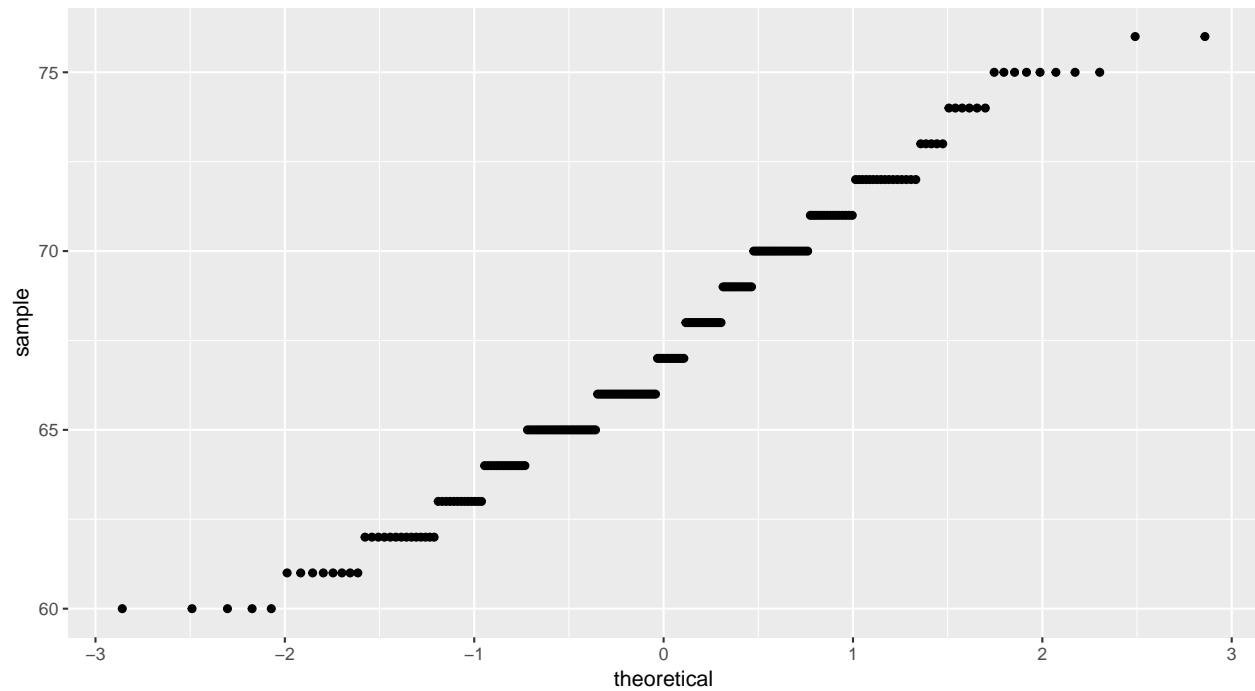
```
# Boxplot  
ggplot(singer, aes(x= "Height", y = height)) + geom_boxplot()
```



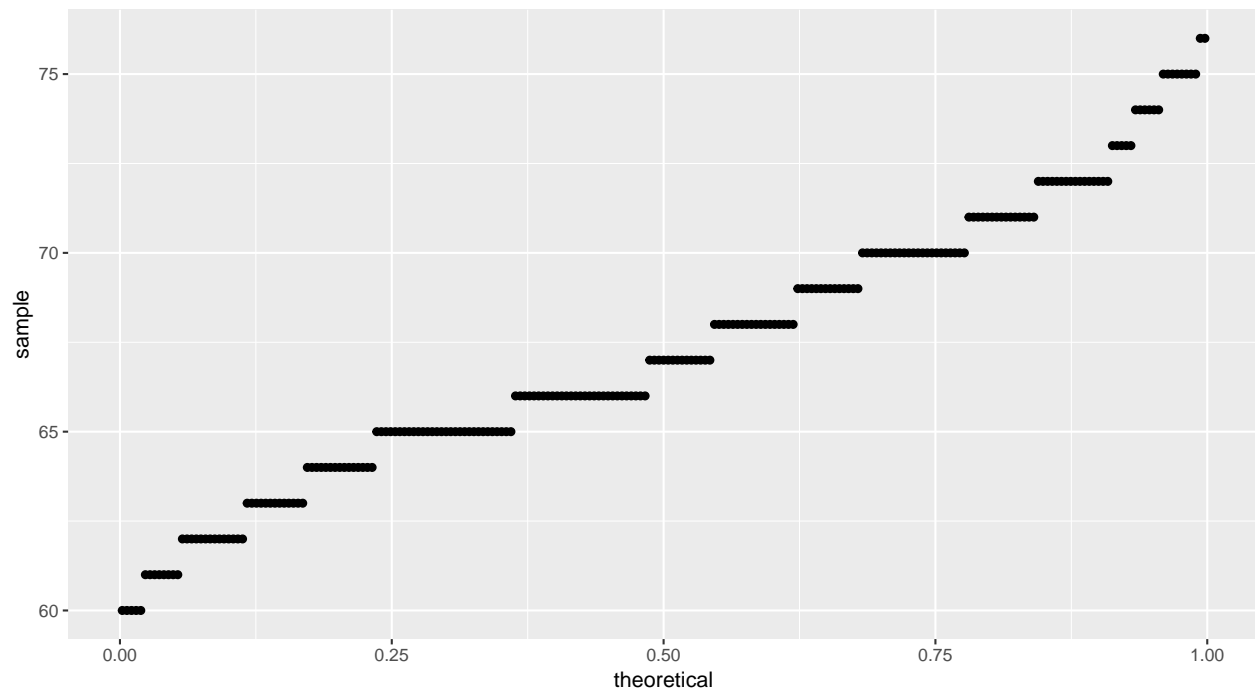
```
# Multiple Boxplots
ggplot(singer, aes(x = voice.part, y = height)) + geom_boxplot()
```



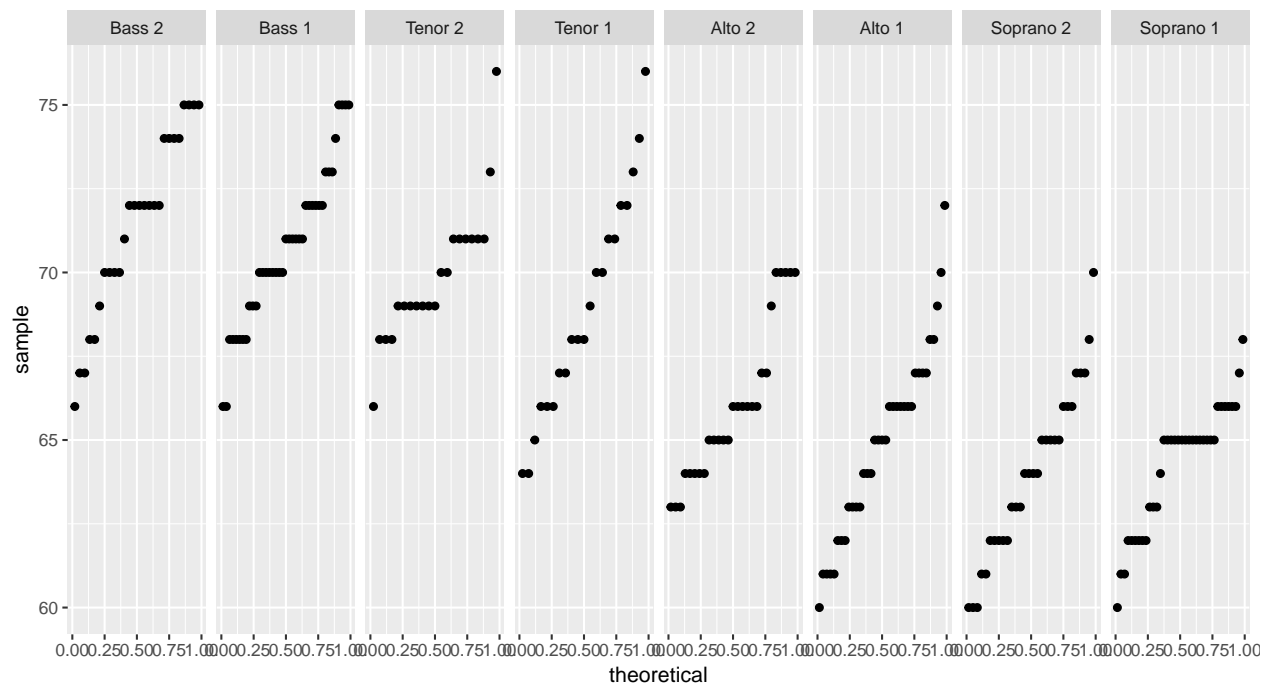
```
# QQ-Plot using a Normal Distribution
ggplot(singer, aes(sample = height)) + stat_qq()
```



```
# QQ-Plot using a Uniform Distribution
ggplot(singer, aes(sample = height)) +
  stat_qq(distribution = qunif)
```



```
# Facet
ggplot(singer, aes(sample = height)) +
  stat_qq(distribution = qunif) +
  facet_grid(~voice.part)
```



```
# The plot above looks cramped, so lets try a grid with columns
# Facet with grid display
ggplot(singer, aes(sample = height)) +
  stat_qq(distribution = qunif) +
  facet_wrap(~voice.part, ncol = 2)
```

