

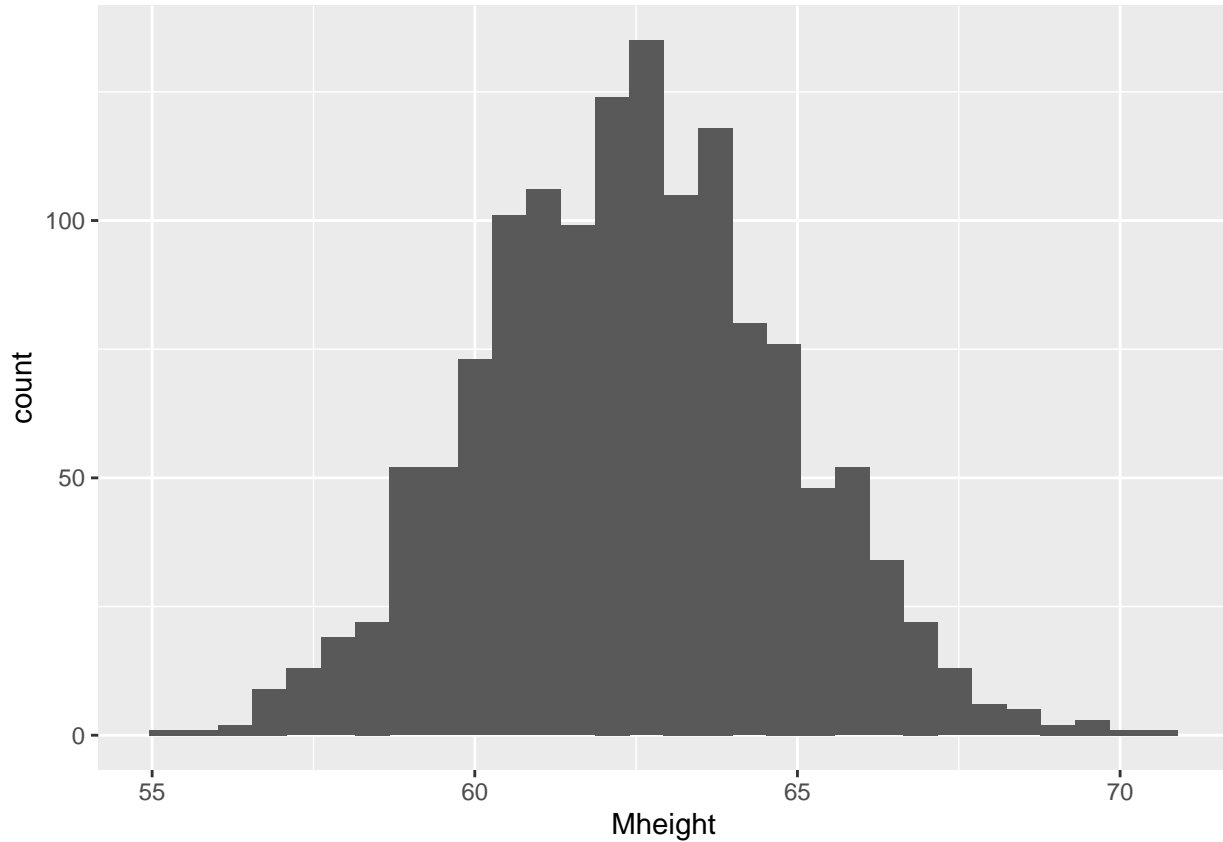
Lecture 5

Jung-Jin Lee

2/4/2020

```
ht <- read.table(file = "heights.txt", header = T, sep = " ")
ggplot(ht, aes(Mheight)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
mean(ht$Mheight)
```

```
## [1] 62.4528
```

```
sd(ht$Mheight)
```

```
## [1] 2.355103
```

```
t.test(ht$Mheight)
```

```
##  
## One Sample t-test  
##  
## data: ht$Mheight  
## t = 983.32, df = 1374, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 62.32821 62.57739  
## sample estimates:  
## mean of x  
## 62.4528
```

```
mu <- 62.2 # null hypothesis  
x <- ht$Mheight  
xbar <- mean(x); S <- sd(x); n <- length(x)  
t <- (xbar - mu) / (S/sqrt(n))  
print(xbar)
```

```
## [1] 62.4528
```

```
print(S)
```

```
## [1] 2.355103
```

```
n
```

```
## [1] 1375
```

```
t
```

```
## [1] 3.980324
```

```
p.value <- 2*(1 - pt(t, df = n-1))
```

```
y <- ht$Dheight  
t.test(y, mu = 63.8)
```

```
##  
## One Sample t-test  
##  
## data: y  
## t = -0.69804, df = 1374, p-value = 0.4853  
## alternative hypothesis: true mean is not equal to 63.8  
## 95 percent confidence interval:  
## 63.6135 63.8886  
## sample estimates:  
## mean of x  
## 63.75105
```

```
t.test(iris$Sepal.Length, mu = 5.1)
```

```
##  
## One Sample t-test  
##  
## data: iris$Sepal.Length  
## t = 10.994, df = 149, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 5.1  
## 95 percent confidence interval:  
## 5.709732 5.976934  
## sample estimates:  
## mean of x  
## 5.843333
```

```
crabs <- read.table("crabs.tsv", header = T, sep = "\t")  
dim(crabs)
```

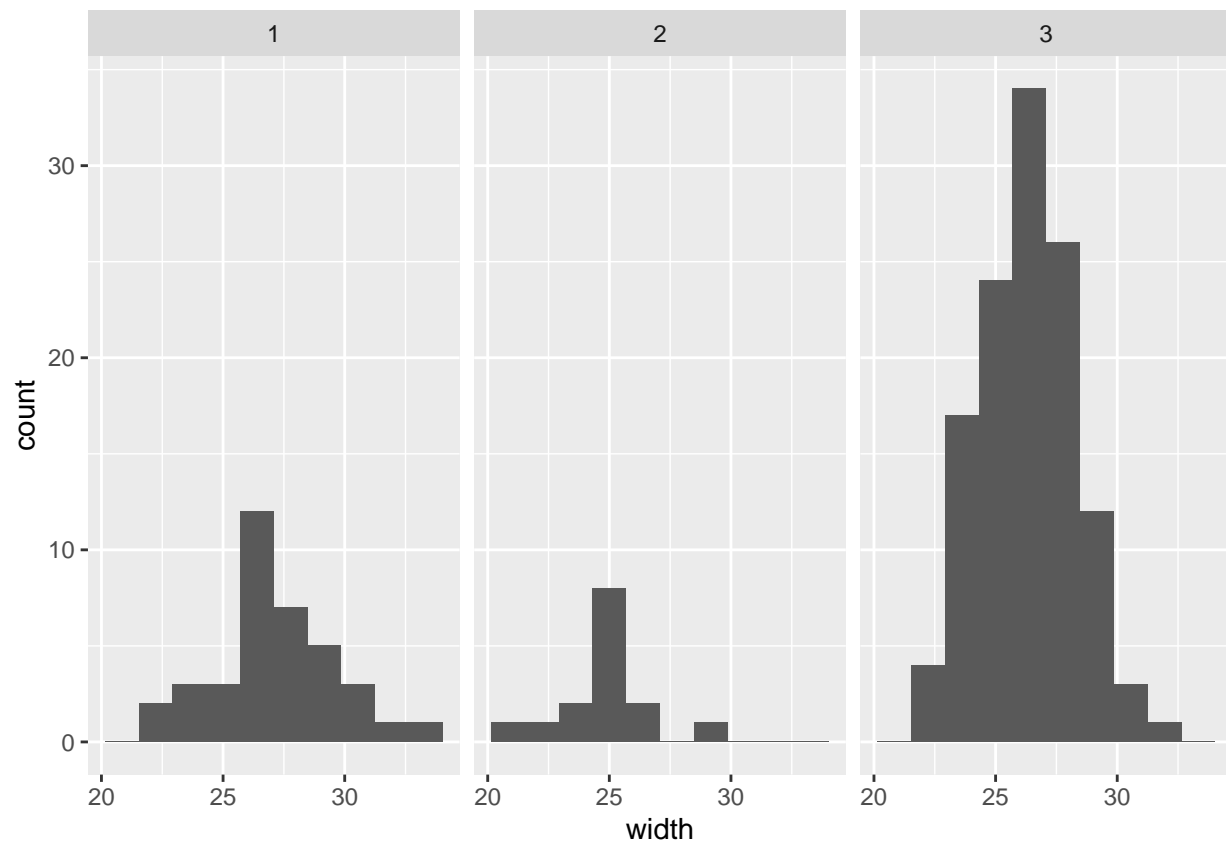
```
## [1] 173 5
```

```
# column names of crabs?  
names(crabs)
```

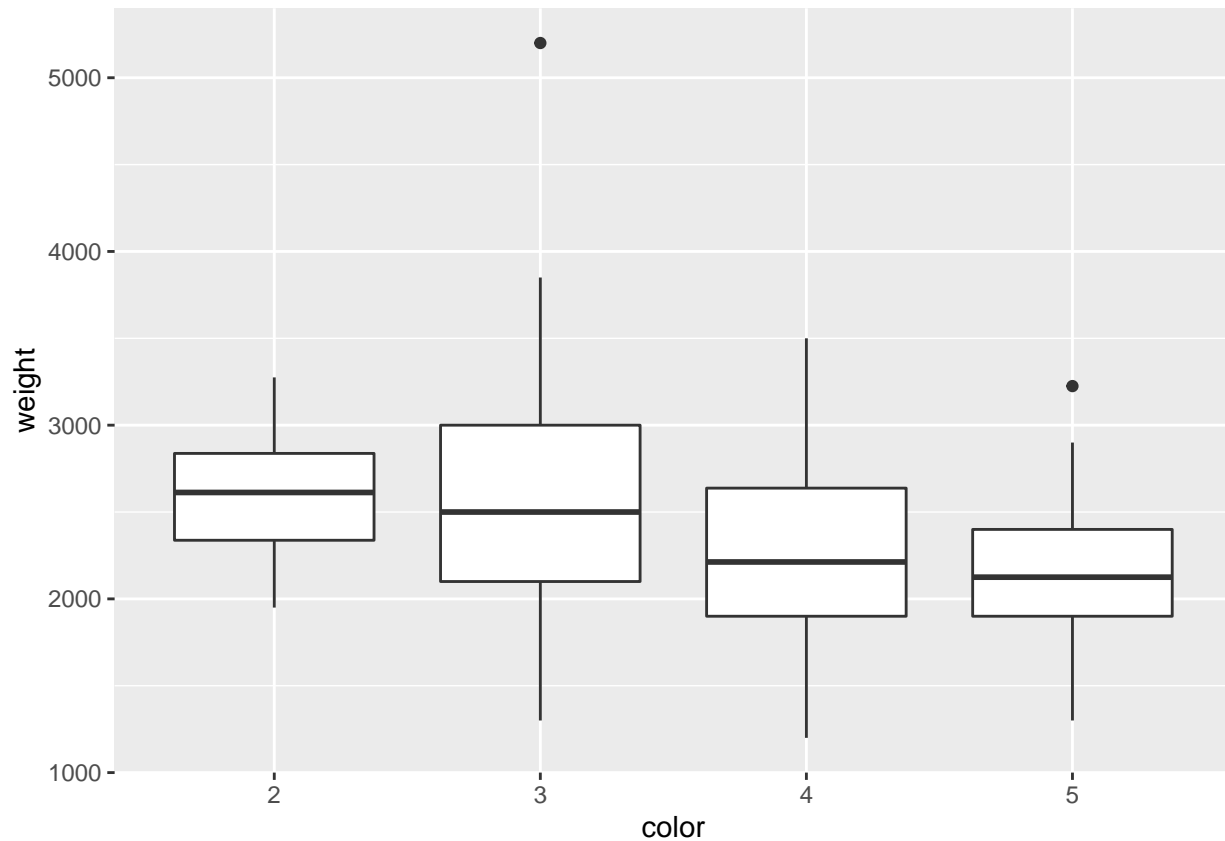
```
## [1] "color" "spine" "width" "satell" "weight"
```

```
crabs <- crabs %>%  
  mutate(color = factor(color)) %>%  
  mutate(spine = factor(spine)) %>%  
  mutate(id = paste0("obs", 1:nrow(crabs)))
```

```
crabs %>%  
  ggplot(aes(width)) +  
  geom_histogram(bins = 10) +  
  facet_wrap(~spine)
```



```
crabs %>%  
  ggplot(aes(color, weight)) +  
  geom_boxplot()
```



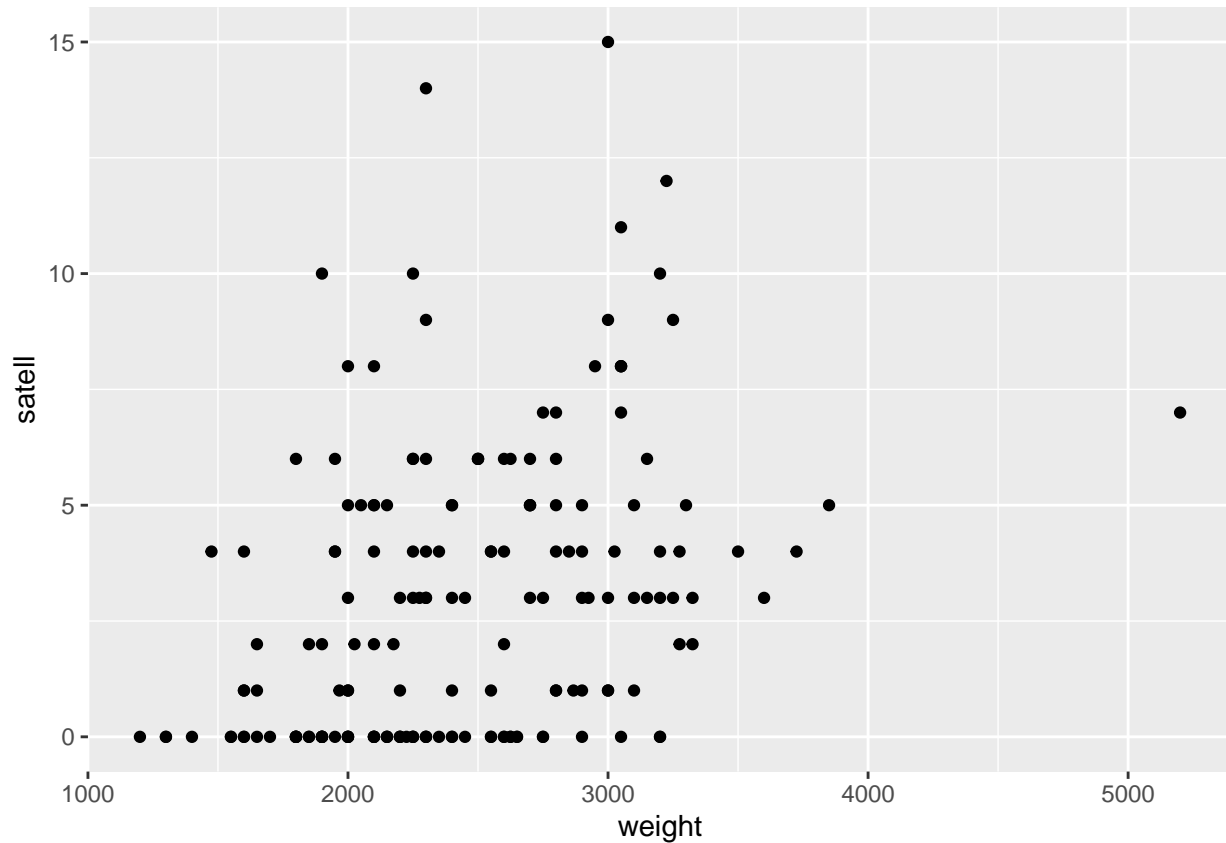
```
crabs %>%
  arrange(desc(weight)) %>%
  head()
```

```
##   color spine width satell weight    id
## 1     3     1  33.5      7  5200 obs141
## 2     3     1  29.7      5  3850 obs91
## 3     3     1  31.7      4  3725 obs147
## 4     3     1  30.3      3  3600 obs50
## 5     4     2  29.8      4  3500 obs160
## 6     3     3  31.9      2  3325 obs115
```

```
crabs %>%
  filter(color == 5) %>%
  arrange(desc(weight)) %>%
  head()
```

```
##   color spine width satell weight    id
## 1     5     3  29.3     12  3225 obs117
## 2     5     3  27.5      0  2900 obs82
## 3     5     1  25.5      0  2750 obs88
## 4     5     3  27.0      0  2625 obs172
## 5     5     3  27.5      0  2600 obs29
## 6     5     3  26.7      0  2450 obs93
```

```
#association between satell and weight
crabs %>%
  ggplot(aes(weight, satell)) +
  geom_point()
```



```
t.test(crabs$width, mu = 26)
```

```
##
##  One Sample t-test
##
## data:  crabs$width
## t = 1.8637, df = 172, p-value = 0.06407
## alternative hypothesis: true mean is not equal to 26
## 95 percent confidence interval:
##  25.98234 26.61535
## sample estimates:
## mean of x
##  26.29884
```

```
t.test(crabs$weight, mu = 2300)
```

```
##
##  One Sample t-test
##
## data:  crabs$weight
```

```
## t = 3.1272, df = 172, p-value = 0.002073
## alternative hypothesis: true mean is not equal to 2300
## 95 percent confidence interval:
##  2350.597 2523.784
## sample estimates:
## mean of x
##  2437.191
```