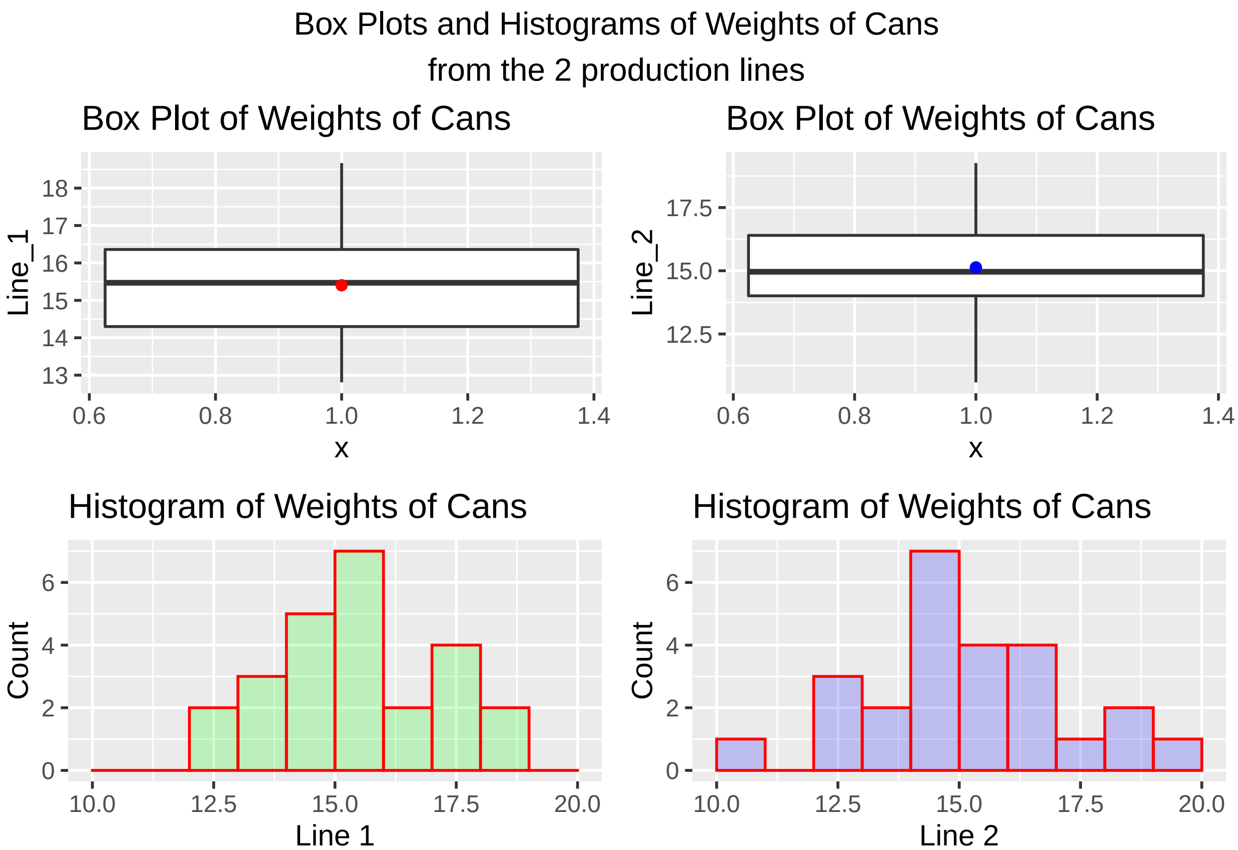
HW1b

Shiloh Bradley

6/9/2020

```r
P <- read.csv("2 production_lines.csv",header=TRUE)

#1(a) Box plots with sample means shown and histograms
B1 <- ggplot(P,aes(x=1,y=Line_1)) + geom_boxplot()+
stat_summary(fun= "mean", geom="point",color="red")+
            labs(title="Box Plot of Weights of Cans")

B2 <- ggplot(P,aes(x=1,y=Line_2)) + geom_boxplot()+
stat_summary(fun = "mean", geom="point",color="blue")+
            labs(title="Box Plot of Weights of Cans")

H1 <- ggplot(data=P, aes(Line_1)) + geom_histogram(breaks=seq(10, 20, by = 1),
                 col="red",
                 fill="green",
                 alpha = .2)+
                 labs(title="Histogram of Weights of Cans") +
                 labs(x="Line 1", y="Count")

H2 <- ggplot(data=P, aes(Line_2)) + geom_histogram(breaks=seq(10, 20, by = 1),
                 col="red",
                 fill="blue",
                 alpha = .2)+
                 labs(title="Histogram of Weights of Cans") +
                 labs(x="Line 2", y="Count")

grid.arrange(B1,B2,H1,H2,ncol=2, top="Box Plots and Histograms of Weights of Cans\nfrom the 2 productio
```




```r
#1(b) Basis stats and 2-sample t-test
summary(P)
```

```
##       Line_1          Line_2
##  Min.   :12.81   Min.   :10.59
##  1st Qu.:14.30   1st Qu.:14.01
##  Median :15.47   Median :14.96
##  Mean   :15.40   Mean   :15.13
##  3rd Qu.:16.36   3rd Qu.:16.40
##  Max.   :18.67   Max.   :19.26
```

```r
#1-sample t-tests
#Test if the mean weights mu1 and mu2 are = 16 oz or not
#Test H0: Mu = 16, vs H1: Mu is not = 16
#Run 1-sample t test on both samples, Line1 and Line2
```

```r
#1(b)
t.test(P$Line_1,mu=16)
```

```
##
##  One Sample t-test
##
## data:  P$Line_1
## t = -1.7494, df = 24, p-value = 0.093
## alternative hypothesis: true mean is not equal to 16
## 95 percent confidence interval:
##  14.69912 16.10728
## sample estimates:
```

```
## mean of x
##   15.4032
```

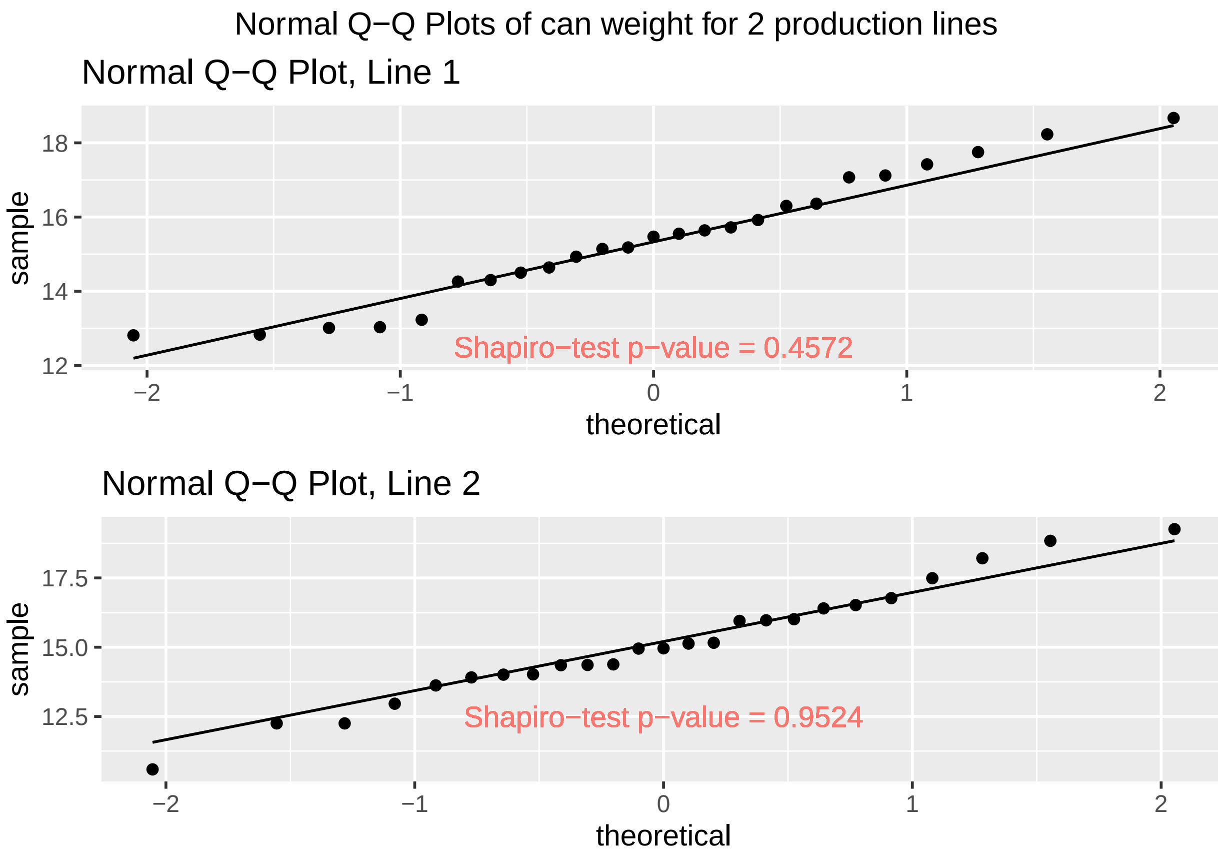
```r
t.test(P$Line_2,mu=16)
```

```
##
##  One Sample t-test
##
## data:  P$Line_2
## t = -2.0804, df = 24, p-value = 0.04833
## alternative hypothesis: true mean is not equal to 16
## 95 percent confidence interval:
##  14.27248 15.99312
## sample estimates:
## mean of x
##   15.1328
```

```r
#1(c)
t.test(P$Line_1,P$Line_2,var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  P$Line_1 and P$Line_2
## t = 0.502, df = 46.193, p-value = 0.618
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.813705  1.354505
## sample estimates:
## mean of x mean of y
##   15.4032   15.1328
```

```r
#1(d)
q1 <- ggplot(P)+stat_qq(aes(sample=Line_1)) +
  geom_qq_line(aes(sample=Line_1))+ggtitle("Normal Q-Q Plot, Line 1")+
  geom_text(aes(x=0, y=12.5, color="red", label="Shapiro-test p-value = 0.4572"))+
  theme(legend.position="none")


q2 <- ggplot(P)+stat_qq(aes(sample=Line_2)) +
  geom_qq_line(aes(sample=Line_2))+ggtitle("Normal Q-Q Plot, Line 2")+
  geom_text(aes(x=0, y=12.5, color="red", label="Shapiro-test p-value = 0.9524"))+
  theme(legend.position="none")

grid.arrange(q1,q2,nrow=2, top="Normal Q-Q Plots of can weight for 2 production lines")
```




```
shapiro.test(P$Line_1) # W = 0.96207, p-value = 0.4572
```

```
##
##  Shapiro-Wilk normality test
##
## data:  P$Line_1
## W = 0.96207, p-value = 0.4572
```

```
shapiro.test(P$Line_2) # W = 0.98409, p-value = 0.9524
```

```
##
##  Shapiro-Wilk normality test
##
## data:  P$Line_2
## W = 0.98409, p-value = 0.9524
```

```
#1(e)Test of Proportions
#proportions of cans with weight < 17
n <- nrow(P)              # 25
x1 <- sum(P$Line_1 < 17)  # 19
x2 <- sum(P$Line_2 < 17)  # 21

prop.test(c(x1,x2),c(n,n),correct=FALSE)  # p-value = 0.4795
```

```
## Warning in prop.test(c(x1, x2), c(n, n), correct = FALSE): Chi-squared
## approximation may be incorrect
```

```
##
##  2-sample test for equality of proportions without continuity
```

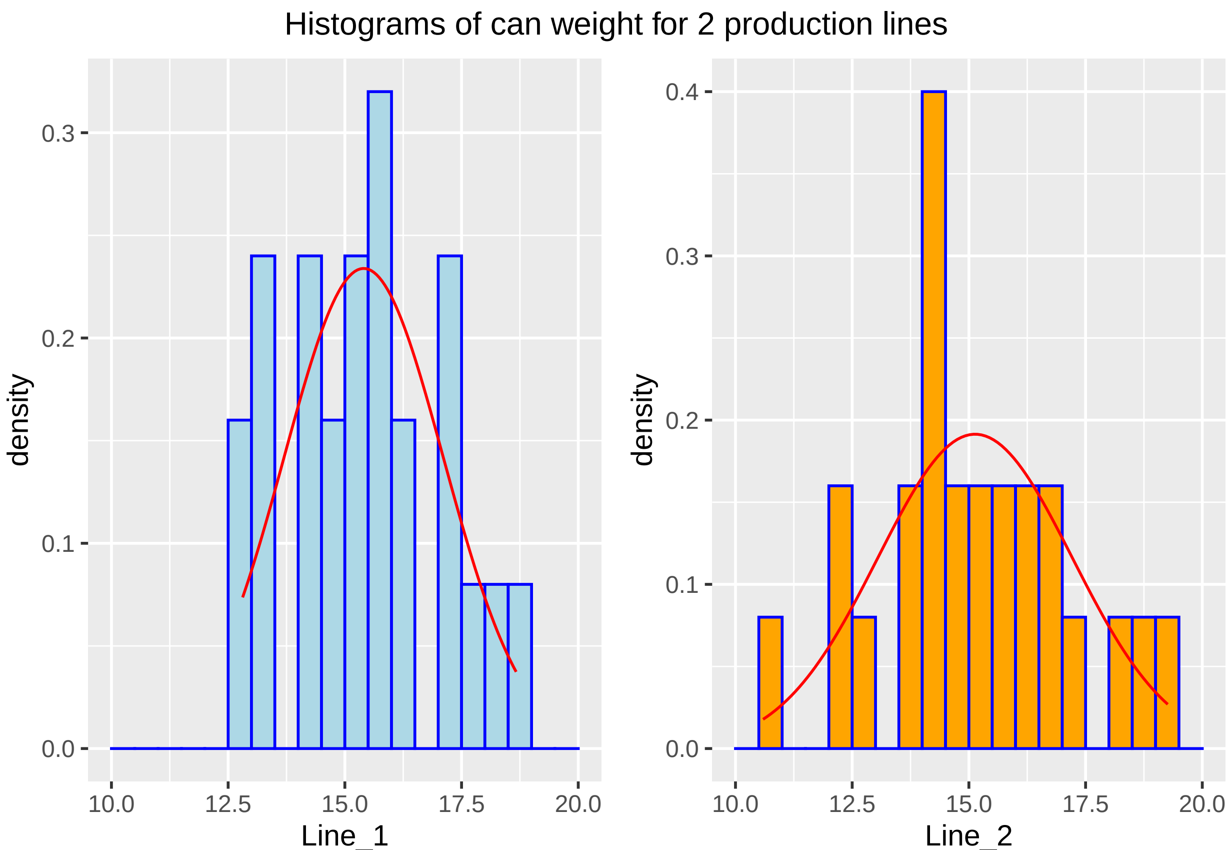
```
##  correction
##
## data:  c(x1, x2) out of c(n, n)
## X-squared = 0.5, df = 1, p-value = 0.4795
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.3006331  0.1406331
## sample estimates:
## prop 1 prop 2
##   0.76   0.84
```

```
prop.test(c(x1,x2),c(n,n),correct=TRUE) # apply continuity correction
```

```
## Warning in prop.test(c(x1, x2), c(n, n), correct = TRUE): Chi-squared
## approximation may be incorrect
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(x1, x2) out of c(n, n)
## X-squared = 0.125, df = 1, p-value = 0.7237
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.3406331  0.1806331
## sample estimates:
## prop 1 prop 2
##   0.76   0.84
```

```
#2(g) verification of normality of samples
h1 <- ggplot(P, aes(x = Line_1)) +
    geom_histogram(aes(y =..density..),
                   breaks = seq(10, 20, by = 0.5),
                   colour = "blue",
                   fill = "lightblue") +
stat_function(fun = dnorm, color = "red", args = list(mean=mean(P$Line_1),sd=sd(P$Line_1)))

h2 <- ggplot(P, aes(x = Line_2)) +
    geom_histogram(aes(y =..density..),
                   breaks = seq(10, 20, by = 0.5),
                   colour = "blue",
                   fill = "orange") +
stat_function(fun = dnorm, color = "red", args = list(mean=mean(P$Line_2),sd=sd(P$Line_2)))

grid.arrange(h1,h2,ncol=2, top="Histograms of can weight for 2 production lines")
```


Histograms of can weight for 2 production lines
density
0.0
0.1
0.2
0.3
10.0
12.5
15.0
17.5
20.0
Line_1
density
0.0
0.1
0.2
0.3
0.4
10.0
12.5
15.0
17.5
20.0
Line_2