Contract Packaging Metrics That Matter: Analysis

Betty Liao

Data Cleaning steps:

1. Load in Data and understad variables' classes.
2. Remove invalid rows.
3. Convert *customer* and *OTIF* to factor variable. Add appropriate label to variable.
4. Convert column 2-5, 8 to date format, **'yyyy/mm/dd H:M'**.
5. Convert *quantity_produced* to common unit - eaches.

Library Used: readr, ggplot2, plyr

Descriptive Analysis on the Data

```
## Parsed with column specification:
## cols(
##   job.id = col_integer(),
##   purchase.order.received.date = col_character(),
##   materials.availablity.date = col_character(),
##   production.started.date = col_character(),
##   production.completed.date = col_character(),
##   quantity.produced = col_integer(),
##   unit.of.measure = col_character(),
##   shipment.shipped.date = col_character(),
##   OTIF = col_integer(),
##   customer = col_character(),
##   `Error check` = col_integer()
## )
```

```
## [1] "Frequency and Proportion of customer"
```

```
##
##         Unilever Procter & Gamble
##              562              553
```

```
##
##         Unilever Procter & Gamble
##        0.5040359        0.4959641
```

```
## [1] "Frequency and Proportion of OTIF"
```

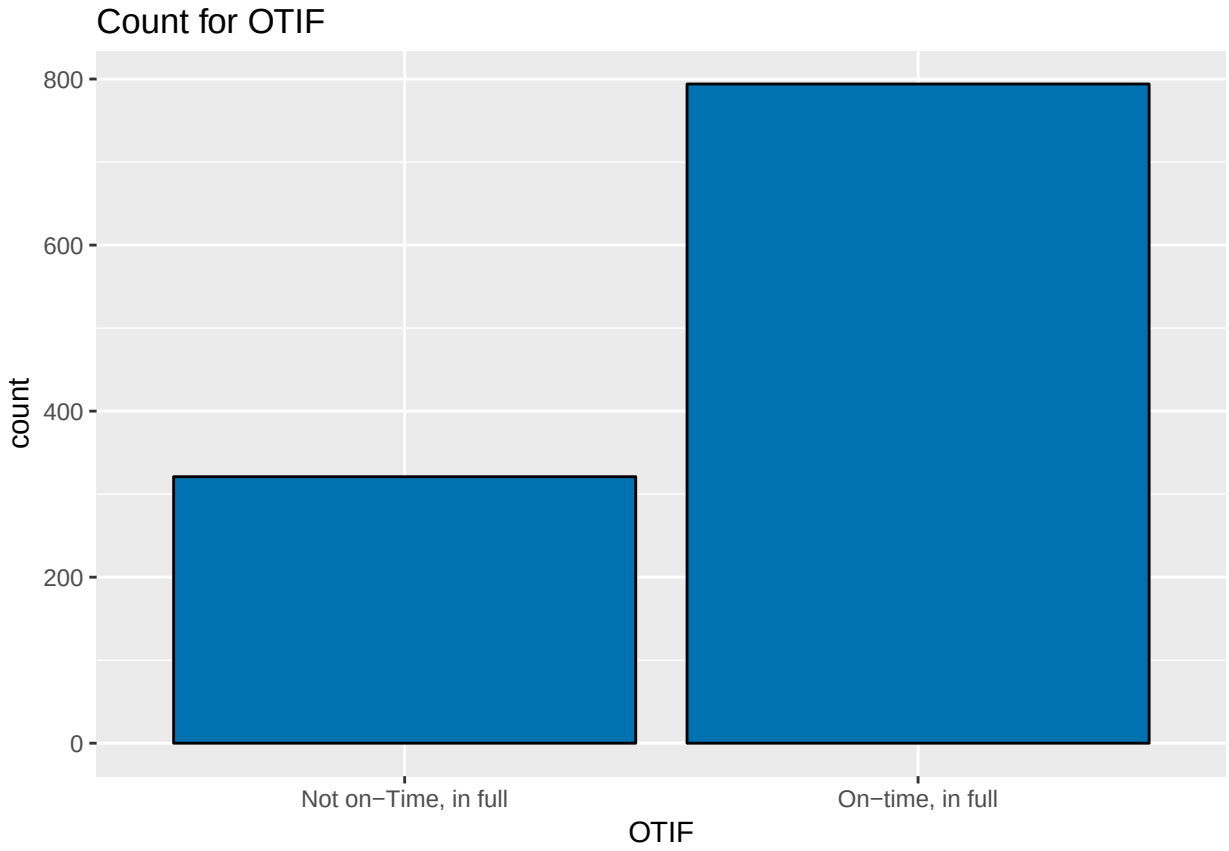
```
##
## Not on-Time, in full     On-time, in full
##                  321                  794
```

```
##
## Not on-Time, in full     On-time, in full
##            0.2878924            0.7121076
```

```
## [1] "Proportion of OTIF for each Customer"
```

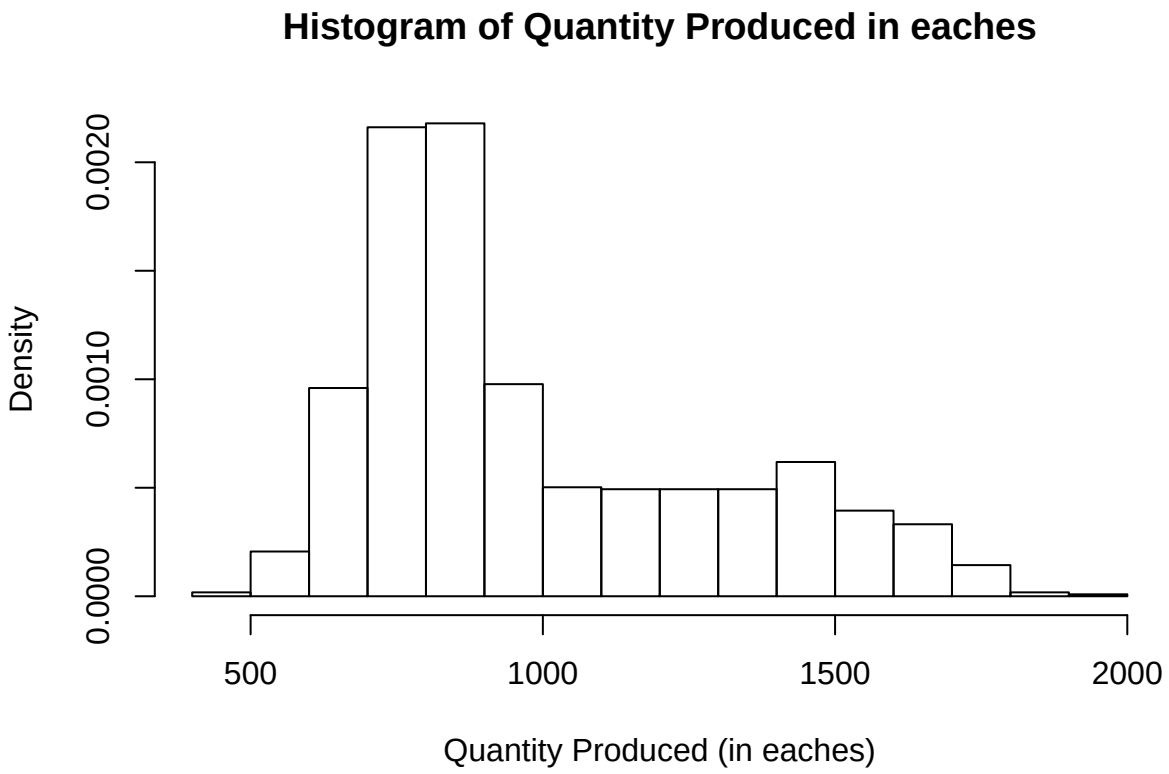
```
##
##                         Unilever Procter & Gamble
```

```
##   Not on-Time, in full 0.2669039        0.3092224
##   On-time, in full     0.7330961        0.6907776
```



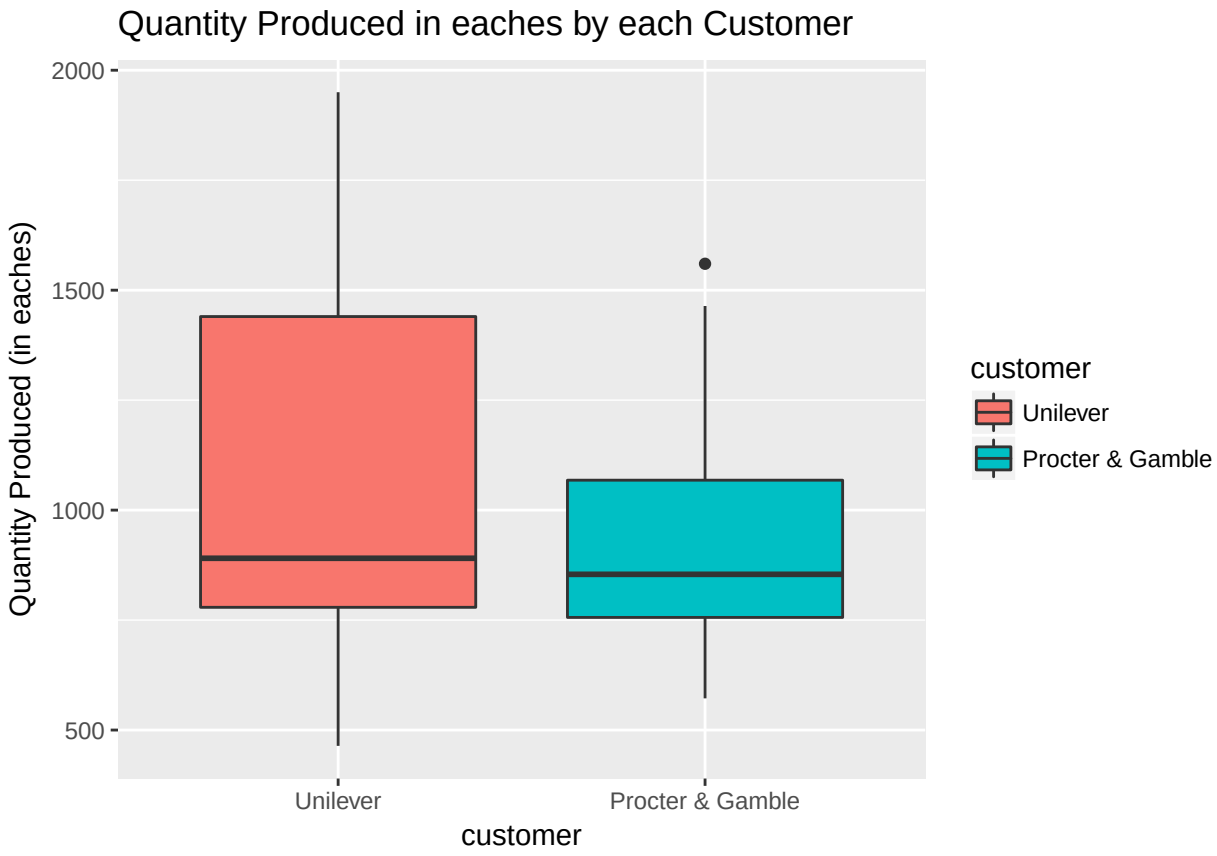

```
## [1] "Summary of quantity produced"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   464.0   764.0   876.0   989.3  1206.0  1950.0
```



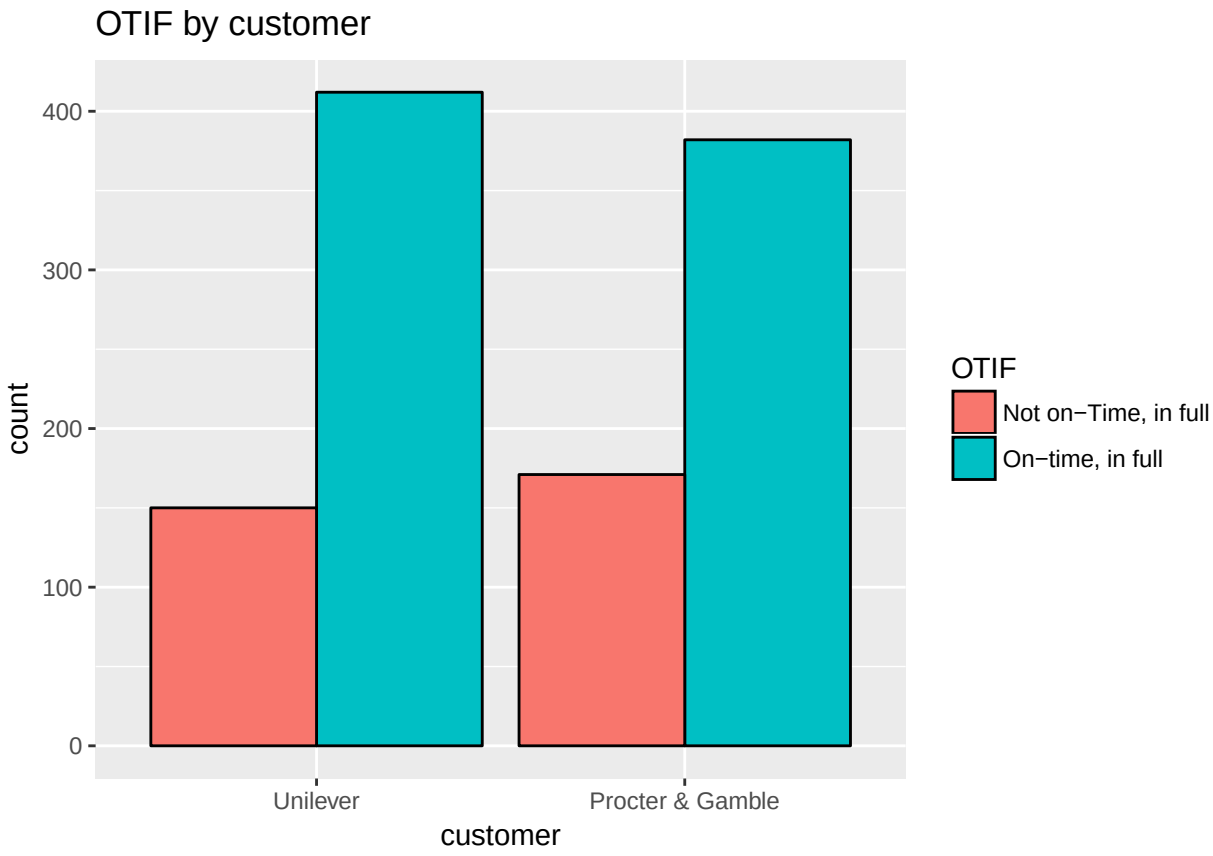

```
## [1] "Mean quantity by customer"
```

```
##            Group.1         x
## 1          Unilever 1065.5961
## 2 Procter & Gamble  911.7993
```



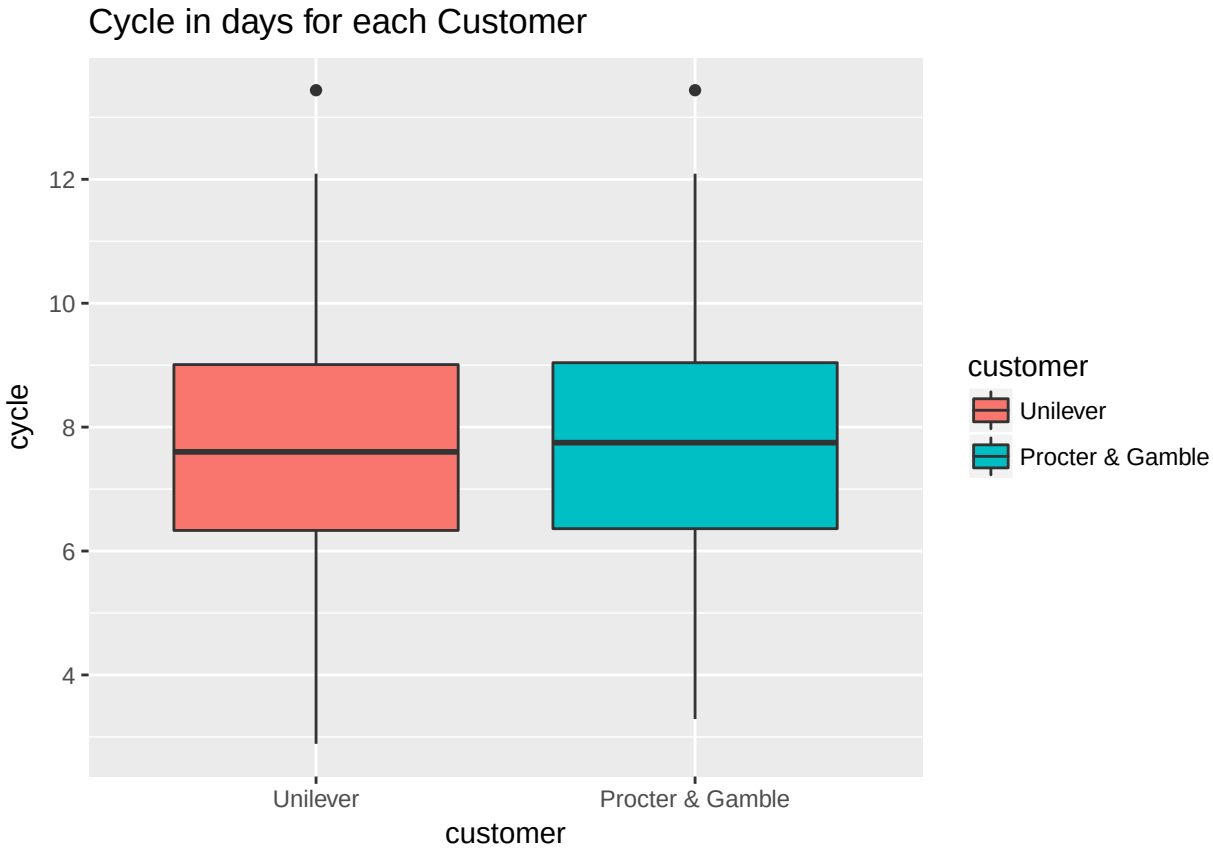

```
## [1] "OTIF frequency by Customer"
```

```
##                   OTIF         customer freq
## 1 Not on-Time, in full         Unilever  150
## 2 Not on-Time, in full Procter & Gamble  171
## 3     On-time, in full         Unilever  412
## 4     On-time, in full Procter & Gamble  382
```




```
## [1] "Average job cycle in days by customer"
```

```
##            Group.1        x
## 1         Unilever 7.647594
## 2 Procter & Gamble 7.676664
```




```
## [1] "Production for 1 eaches in minutes"
```

```
##            Group.1        x
## 1         Unilever 2.355942
## 2 Procter & Gamble 1.964755
```

Comments:

- Before removing error data, the customer are evenlly distributed, each customer has 600 data points. After removing 86 rows, Unilever has 562 order and P&G has 553 order. The proportion is 50.4% and 49.6%.
- 28.78% of time has a result of OTIF = 0 and 71.22% of time has a result of OTIF = 1.
- After converting all the units to eaches, the mean of quantity produced is approximately 989 eaches and the median is 876 eaches. With Unilever producing an average of 1066 eaches and P&G producing an average of 912 eaches.
- Breaking down OTIF frequency by customer, we see that Unilever has more on-time package than P&G by 30 time (412 vs 382). In addition, Unilver has less not on-time package than P&G by 21 time (150 vs 171). You can visualize the difference by the above bar graph.
- Average job cycle from receiving PO to shipment are very close. Looking at box-plot we can expect no difference between average job cycle.
- Production efficiency (per eaches) for each customer: Unilever takes approximately 2.36 minutes to produce 1 eaches and P&G takes approximately 1.96 minutes to produce 1 eaches.

Question 1: What is the average shift length?

```
## Time difference of 9.869925 mins
```

```
## [1] "Average shift length (in minutes) by customer"
```

```
##            Group.1         x
## 1         Unilever 9.937040
## 2 Procter & Gamble 9.801718
```

Comments:

The average shift length is approximately 9.87 hours. If we seperate according to customers, Unilever has an average of 9.93 hours and P&G has an average of 9.8 hours, with P&G being slightly shorter than Unilever.

Question 2:

(a) What is the change in probability of OTIF 3 days after receiving the PO vs 4 days? (b) How many days can the supplier afford to wait after receiving the PO to start production if they hope to be OTIF?

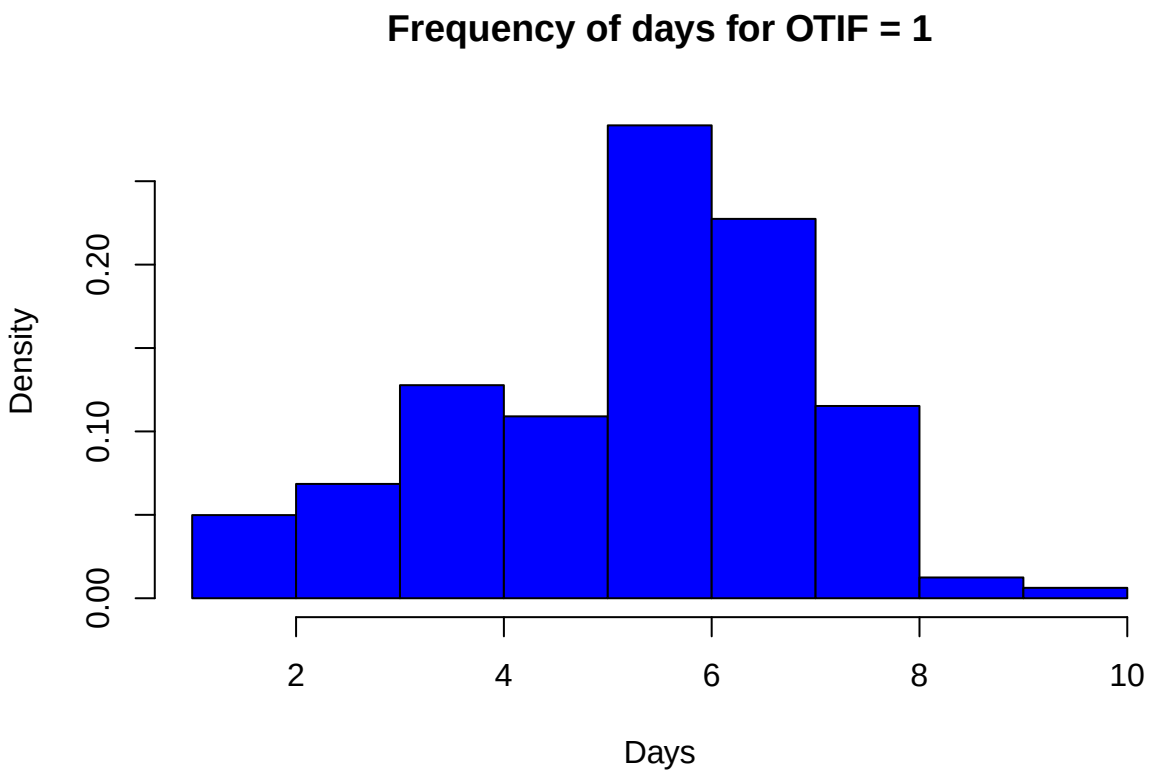
```
## Parsed with column specification:
## cols(
##   job.id = col_integer(),
##   purchase.order.received.date = col_character(),
##   materials.availablity.date = col_character(),
##   production.started.date = col_character(),
##   production.completed.date = col_character(),
##   quantity.produced = col_integer(),
##   unit.of.measure = col_character(),
##   shipment.shipped.date = col_character(),
##   OTIF = col_integer(),
##   customer = col_character(),
##   `Error check` = col_integer()
## )
```

```
## [1] "Count OTIF by 3 days"
```

```
##   OTIF three_days freq
## 1    0          0  316
## 2    0          1    5
## 3    1          0  761
## 4    1          1   33
```

```
## [1] "Count OTIF by 4 days"
```

```
##   OTIF four_days freq
## 1    0         0  306
## 2    0         1   15
## 3    1         0  744
## 4    1         1   50
```




```
## [1] "Day by OTIF - Mean"
```

```
##                 Group.1        x
## 1     On-Time, in full 5.735202
## 2 Not on-time, in full 4.221662
```

```
## [1] "Day by OTIF - Median"
```

```
##                 Group.1 x
## 1     On-Time, in full 6
## 2 Not on-time, in full 4
```

Comments:

- Total OTIF=1 for four_days = 1 is 50. Thus, probability of OTIF 4 days after PO is 50/65 = 0.77
- Total OTIF=1 for three_days = 1 is 33. Thus, probability of OTIF 3 days after PO is 33/38 = 0.87
- a.) Change in probability of OTIF: Decrease in 10%
- b.) Assume "affort to wait" mean obtaining OTIF = 1 for at least half of the time, then we will take the median of 6 days. (Ans: 6 days or less)

Question 3: Is the difference in quantity produced between P&G and Unilever statistically significant?

```
## Parsed with column specification:
## cols(
##   job.id = col_integer(),
##   purchase.order.received.date = col_character(),
##   materials.availablity.date = col_character(),
##   production.started.date = col_character(),
##   production.completed.date = col_character(),
##   quantity.produced = col_integer(),
##   unit.of.measure = col_character(),
##   shipment.shipped.date = col_character(),
##   OTIF = col_integer(),
##   customer = col_character(),
##   `Error check` = col_integer()
## )
```

```
## [1] "Test to compare equal variance: result indicates variances are not equal"
```

```
##
##  F test to compare two variances
##
## data:  copackag$quant_each[copackag$customer == 1] and copackag$quant_each[copackag$customer == 2]
## F = 1.0488, num df = 561, denom df = 552, p-value = 0.5744
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.8880352 1.2385511
## sample estimates:
## ratio of variances
##            1.04881
```

```
##
##  Welch Two Sample t-test
##
## data:  copackag$quant_each[copackag$customer == 1] and copackag$quant_each[copackag$customer == 2]
## t = -2.4786, df = 1112.9, p-value = 0.01334
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -80.443568  -9.355914
## sample estimates:
## mean of x mean of y
##  428.6174  473.5172
```

Ans: No

Comments:

The null hypothesis assume we have equal mean of quantity produced for the two customers. However, we have a p-value of 0.01334 which is less than 0.05. Thus we reject the null hypothesis and conclude the mean quanitity produced between the two customers are not equal.

Question 4: Assuming everything else is constant, what is the probability of hitting OTIF if the customer was P&G?

```
## Parsed with column specification:
## cols(
##   job.id = col_integer(),
##   purchase.order.received.date = col_character(),
##   materials.availablity.date = col_character(),
##   production.started.date = col_character(),
##   production.completed.date = col_character(),
##   quantity.produced = col_integer(),
##   unit.of.measure = col_character(),
##   shipment.shipped.date = col_character(),
##   OTIF = col_integer(),
##   customer = col_character(),
##   `Error check` = col_integer()
## )
```

```
## [1] "Count frequency of OTIF by customer"
```

```
##                   OTIF         customer freq
## 1     On-Time, in full         Unilever  150
## 2     On-Time, in full Procter & Gamble  171
## 3 Not on-time, in full         Unilever  412
## 4 Not on-time, in full Procter & Gamble  382
```

Ans: Probability will be OTIF/(total OTIF) = 171/553 = 31% (0.30922)

Apendix: R code

```
# R code for Data Analysis
## Data Loading and Cleansing
library(readr)
copackag <- read_csv("~/Desktop/Case study/DA Assignment 2018_Nulogy/copackager_table.csv")

# Remove Error rows
copackag <- copackag[!(copackag$`Error check` == 1),]
# Define variables type and add labels
copackag$customer[copackag$customer == "Unilever"] <- 1
copackag$customer[copackag$customer == "Procter & Gamble"] <- 2
copackag$customer <- factor(copackag$customer, levels = c(1,2),
                            labels = c("Unilever", "Procter & Gamble"))
copackag$OTIF <- factor(copackag$OTIF, levels = c(0,1),
                        labels = c("Not on-Time, in full", "On-time, in full"))

# convert date info in format 'yy/mm/dd H:M'
dates <- c(2:5, 8)  # Specify Date columns
copackag[,dates] <- sapply(copackag[,dates], as.POSIXlt, "%Y-%m-%d %H:%M",tz=Sys.timezone())

# convert quantity to common unit - eaches
copackag$quant_each <-
  ifelse((copackag$customer == "Procter & Gamble") &
           (copackag$unit.of.measure == "cases"), copackag$quantity.produced*2,
  ifelse((copackag$customer == "Procter & Gamble") &
           (copackag$unit.of.measure == "pallets"), copackag$quantity.produced*12,
  ifelse((copackag$customer == "Unilever") & (copackag$unit.of.measure == "cases"),
         copackag$quantity.produced*2,
  ifelse((copackag$customer == "Unilever") & (copackag$unit.of.measure == "pallets"),
         copackag$quantity.produced*15, copackag$quantity.produced))))

# Display first 20 rows of data
head(copackag, n=20)
# list types for each attribute
sapply(copackag, class)

## Descriptive Analysis
library(ggplot2)
# Frequency and Proportion of customer
table(copackag$customer); prop.table(table(copackag$customer))

# Frequency and Proportion of OTIF
table(copackag$OTIF); prop.table(table(copackag$OTIF))
# Proportion of OTIF for customer
prop.table(table(copackag$OTIF, copackag$customer), margin = 2)
# Visualize quantity produce by OTIF
ggplot(data=copackag, aes(x = `OTIF`)) + geom_bar(colour="black", fill="#0072B2") +
  ggtitle("Count for OTIF")

# Frequency and Proportion of quantity produced
summary(copackag$quant_each)
hist(copackag$quant_each, main = "Histogram of Quantity Produced in eaches",
     xlab="Quantity Produced (in eaches)", prob=TRUE)

# Mean quantity by customer
aggregate(copackag$quant_each, by=list(copackag$customer), FUN=mean)
ggplot(copackag, aes(x = customer, y = quant_each, fill = customer)) + geom_boxplot() +
```

```
  ggtitle("Quantity Produced in eaches by each Customer") +
  ylab("Quantity Produced (in eaches)")

# Visualize OTIF by customer
library(plyr)
count(copackag, c("OTIF", "customer"))
ggplot(data=copackag, aes(customer, ..count..)) + geom_bar(aes(fill = OTIF), colour="black",
      position = "dodge") + ggtitle("OTIF by customer")

# Average job cycle by customer
copackag$cycle <- as.numeric(copackag$shipment.shipped.date - copackag$purchase.order.received.date)
aggregate(copackag$cycle, by=list(copackag$customer), FUN=mean)
ggplot(copackag, aes(x = customer, y = cycle, fill = customer)) + geom_boxplot() +
  ggtitle("Cycle in days for each Customer")

# Average production time for each customer by 1 eaches
copackag$pro_time <- as.numeric(copackag$production.completed.date -
                                  copackag$production.started.date)
copackag$pro_eff <- copackag$quant_each/copackag$pro_time
aggregate(copackag$pro_eff, by=list(copackag$customer), FUN=mean)

### Assignment Codes:
## Q1
# Calculate Shift Length in minutes
copackag$shift_len <- copackag$production.completed.date - copackag$production.started.date
mean_shiftlen <- mean(copackag$shift_len)/60; mean_shiftlen
# Average shift length by customer
copackag$shift_len_min <- (copackag$shift_len)/60
aggregate(copackag$shift_len_min, by=list(copackag$customer), FUN=mean)

## Q2
library(plyr)
## Probability of OTIF 3 days after PO
# Calculate time in days from receiving PO to shipment
copackag$len <- as.numeric(copackag$shipment.shipped.date - copackag$purchase.order.received.date)
copackag$len <- floor(copackag$len)
# Dummy coding for 3 days
copackag$three_days <- ifelse((copackag$len == 3), 1, 0)
# Count OTIF by 3 days
count(copackag, c("OTIF", "three_days"))

## Probability of OTIF 4 days after PO
# Dummy coding for 4 days
copackag$four_days <- ifelse((copackag$len == 4), 1, 0)
# Count OTIF by 4 days
count(copackag, c("OTIF", "four_days"))

## Difference between Production start date and receiving PO
copackag$day <- as.numeric(copackag$production.started.date - copackag$purchase.order.received.date)
copackag$day <- copackag$day/24
copackag$day <- floor(copackag$day)  # Round down day to integer
copackag$day <- as.integer(copackag$day)
# Histogram of OTIF = 1 and day
library(ggplot2)
copackag$OTIF <- factor(copackag$OTIF, levels = c(0,1),
                        labels = c("On-Time, in full", "Not on-time, in full"))
hist(copackag$day[copackag$OTIF=="On-Time, in full"], main = "Frequency of days for OTIF = 1",
     xlab="Days", col="blue", prob=TRUE)
```

```r
# Find mean and median of day for OTIF = 1
aggregate(copackag$day, by=list(copackag$OTIF), FUN=mean)
aggregate(copackag$day, by=list(copackag$OTIF), FUN=median)

## Q3
## Test to compare equal variance: result indicates variances are not equal
var.test(copackag$quant_each[copackag$customer==1], y = copackag$quant_each[copackag$customer==2],
         alternative = "two.sided")
# Results indicate variance are not equal

t.test(copackag$quant_each[copackag$customer== 1], y = copackag$quant_each[copackag$customer==2],
       mu = 0, alternative = c("two.sided"), paired = FALSE,var.equal = FALSE)

## Q4
# Count frequency of OTIF by customer
count(copackag, c("OTIF", "customer"))
```