

Untitled

Halid Kopanski

4/12/2022

Problem 1 a)

```
set.seed(1)

theta_0 <- 0
theta_a <- c(0.1, 0.4, 0.7, 1)

alpha <- 0.05

Y1 <- runif(10000, theta_0, theta_0 + 1)
Y2 <- runif(10000, theta_0, theta_0 + 1)

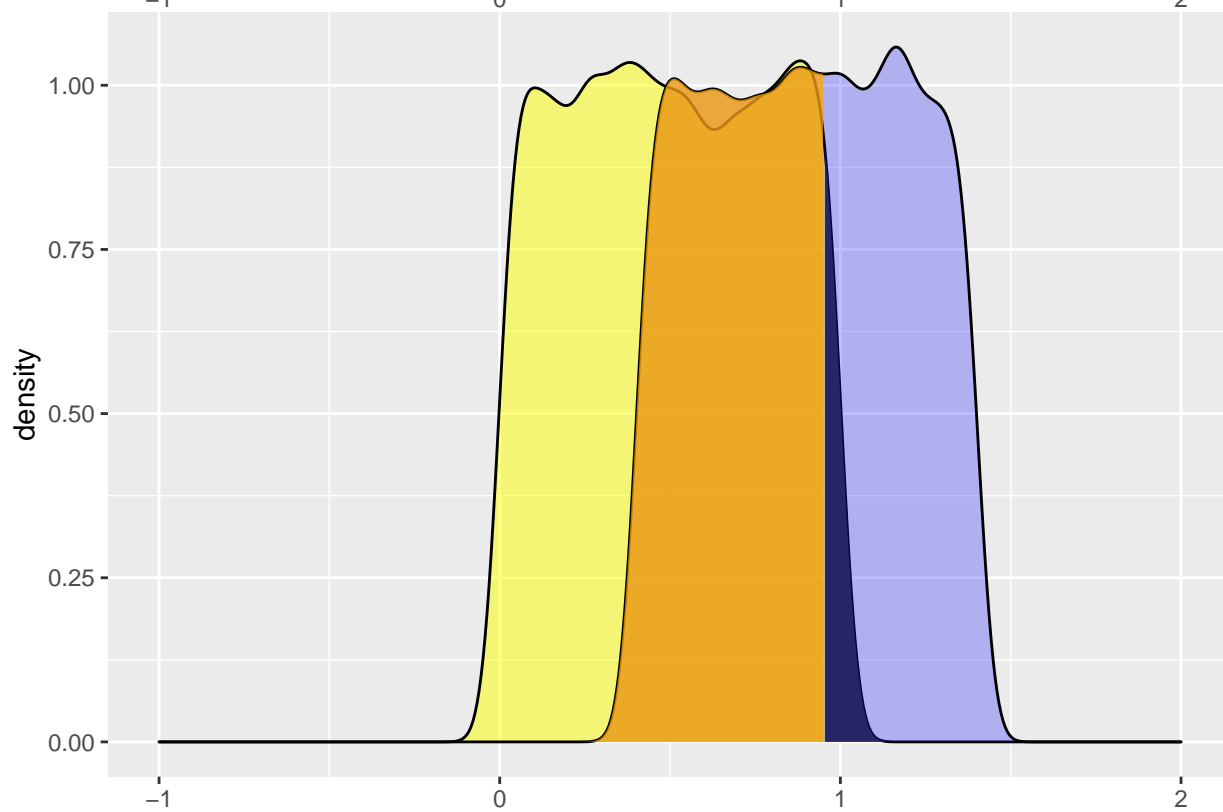
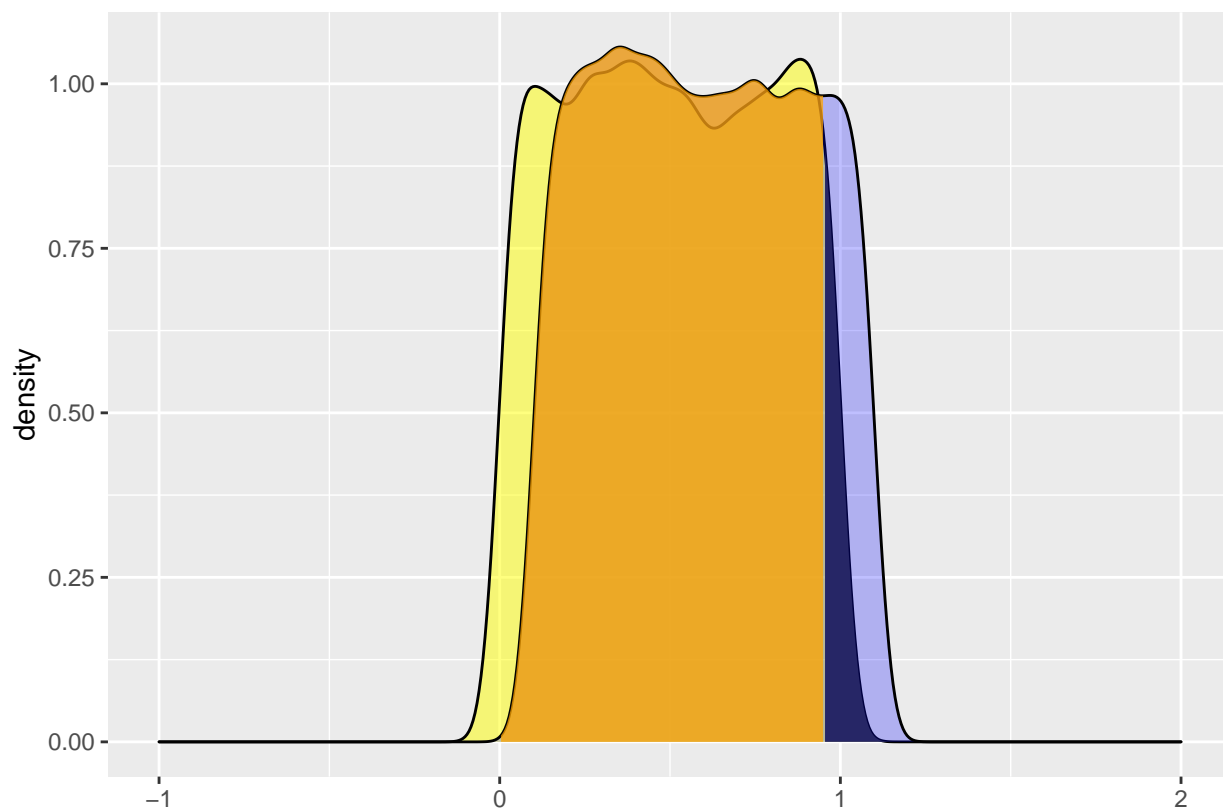
Y3 <- Y1 + Y2

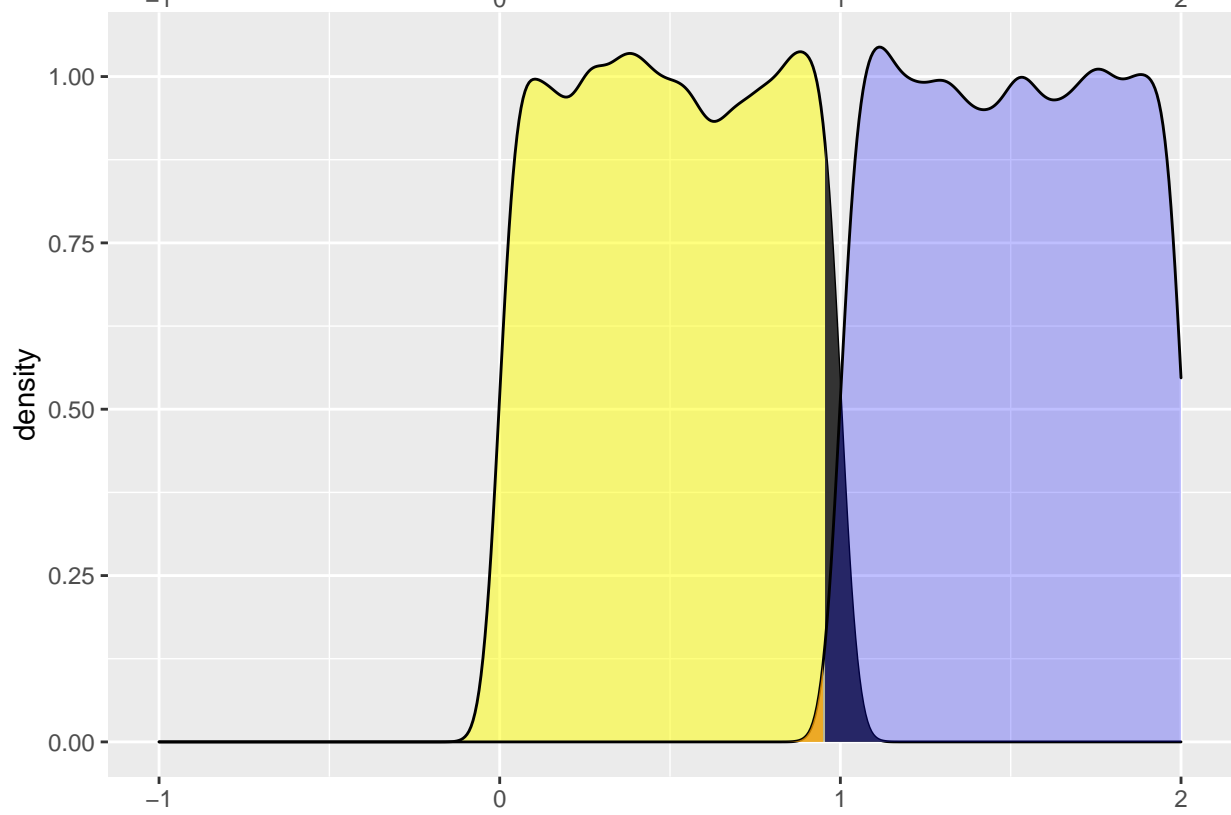
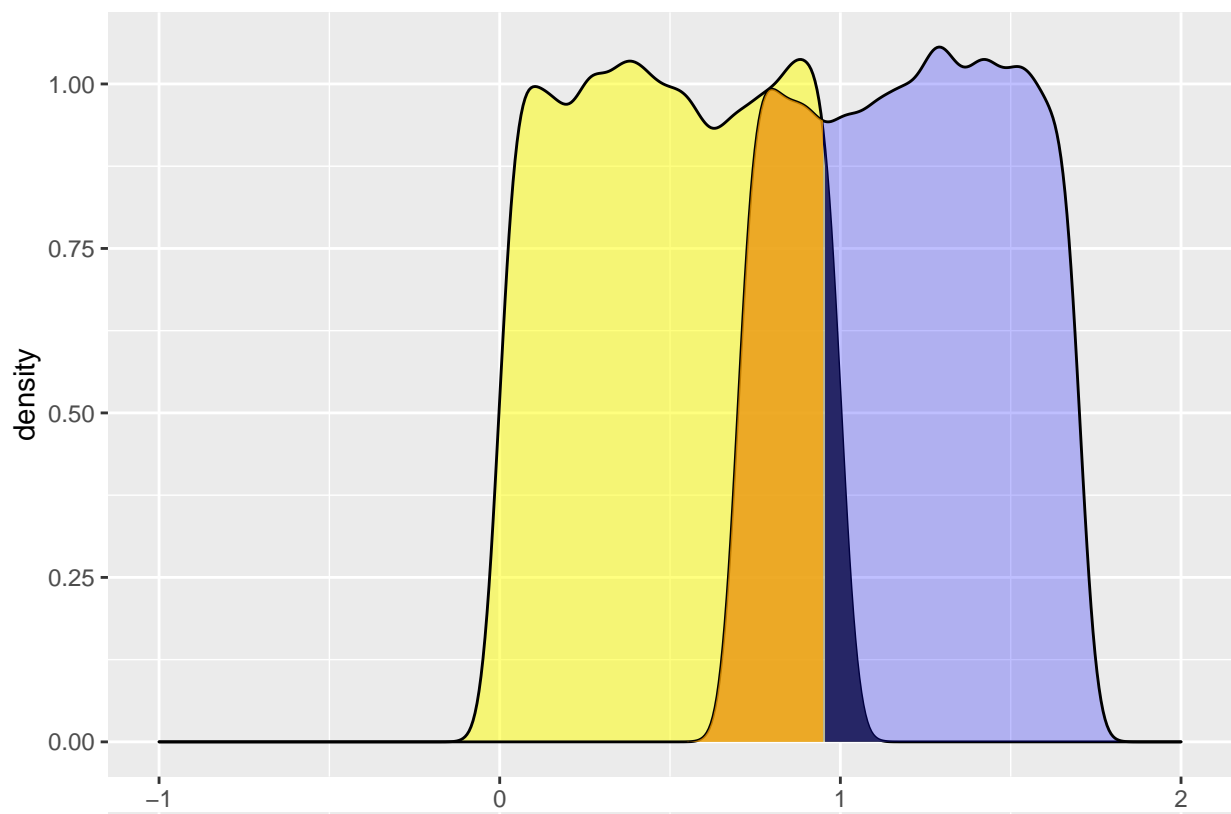
for (i in 1:4){
  YA <- runif(10000, theta_a[i], theta_a[i] + 1)
  df_Y <- tibble(Y1, Y2, Y3, YA)

  plot_ <-
    ggplot(data = df_Y) +
    geom_density(aes(Y1), fill = "yellow", alpha = 0.5) +
    geom_area(
      aes(x = stage(Y1,
                    after_stat = oob_censor(x, (1 - alpha)))),
        stat = "density") +

    geom_density(aes(YA), fill = "blue", alpha = 0.25) +
    geom_area(
      aes(x = stage(YA, after_stat = oob_censor(x, c(0, (1 - alpha))))),
        stat = "density", fill = "orange", alpha = 0.75) + xlab("") + xlim(-1, 2)

  print(plot_)
}
```





Problem 1 b)

```
set.seed(1)

theta_0 <- 0
theta_a <- c(0.1, 0.4, 0.7, 1)

c <- 1.684

Y1 <- runif(10000, theta_0, theta_0 + 1)
Y2 <- runif(10000, theta_0, theta_0 + 1)

Y3 <- Y1 + Y2

for (i in 1:4){
  YA1 <- runif(10000, theta_a[i], theta_a[i] + 1)
  YA2 <- runif(10000, theta_a[i], theta_a[i] + 1)

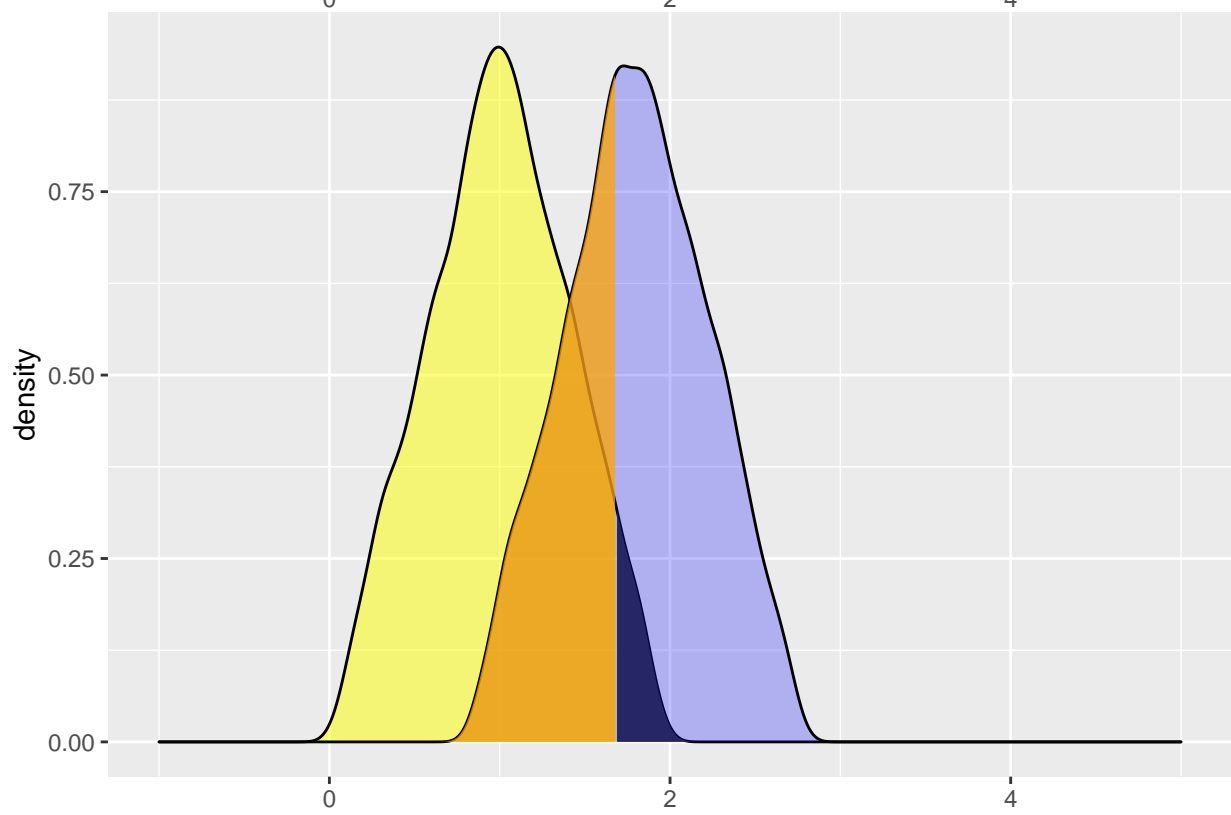
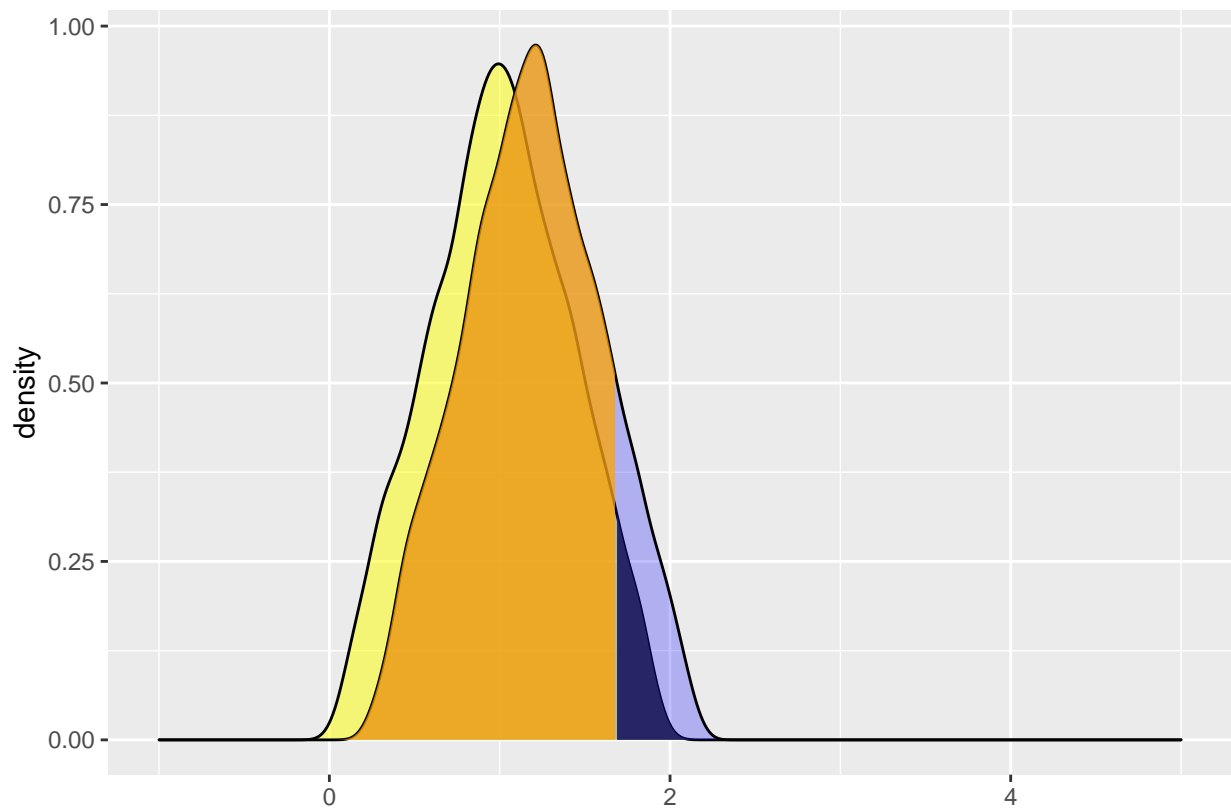
  YA <- YA1 + YA2
  df_Y <- tibble(Y1, Y2, Y3, YA)

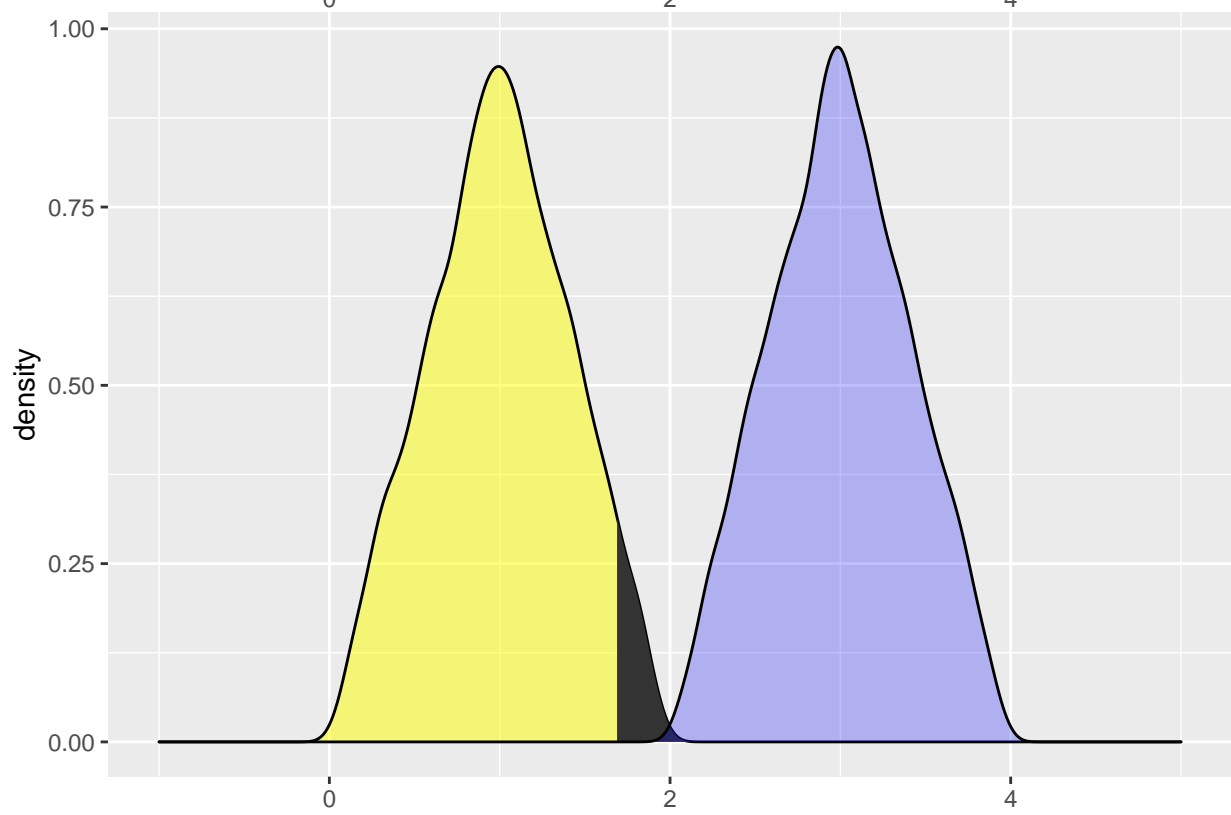
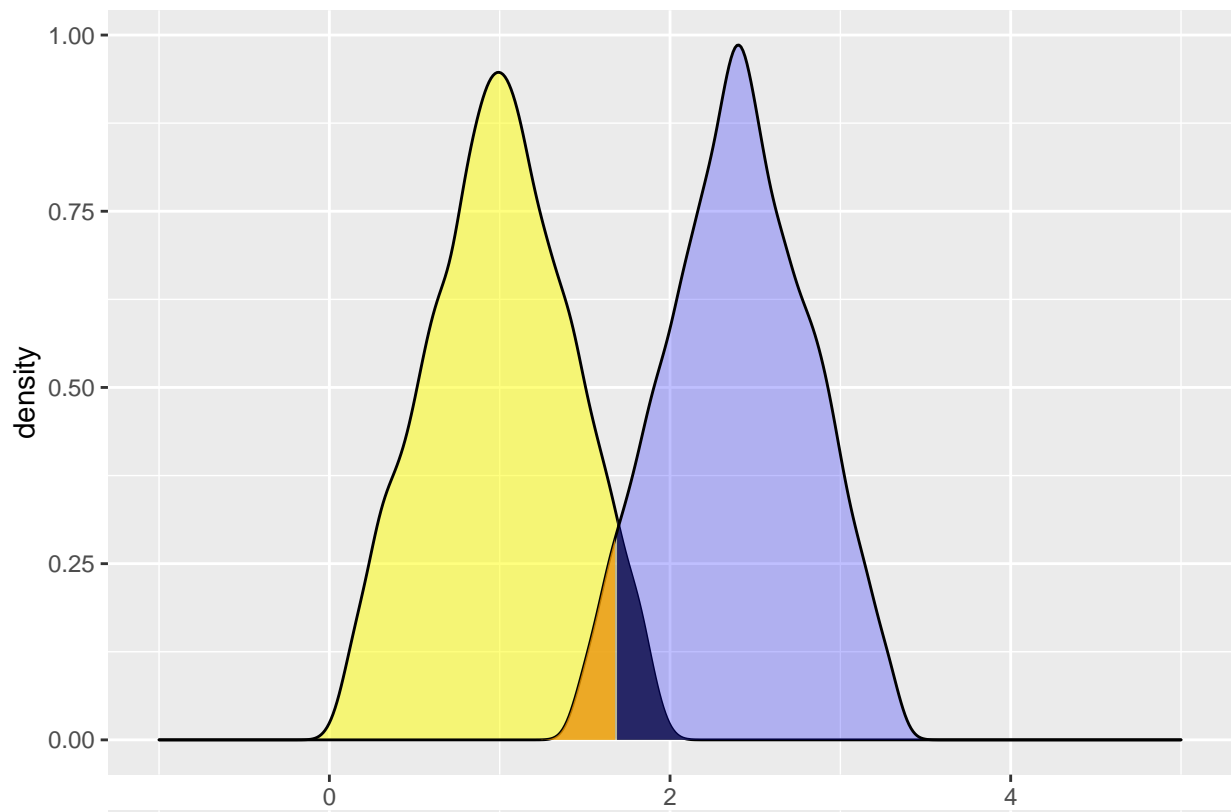
  plot_ <-
    ggplot(data = df_Y) +
    geom_density(aes(Y3), fill = "yellow", alpha = 0.5) +
    geom_area(

    aes(x = stage(Y3,
                  after_stat = oob_censor(x, c))),
        stat = "density") +

    geom_density(aes(YA), fill = "blue", alpha = 0.25) +
    geom_area(
      aes(x = stage(YA, after_stat = oob_censor(x, c(0, c))
                    )
          ),
          stat = "density",
          fill = "orange",
          alpha = 0.75) +
    xlab("") + xlim(-1, 5)

  print(plot_)
}
```





Problem 2

```
no_accidents <- c(1735, 1657, 1812, 1962, 1965, 1849, 1755,
                  1892, 1796, 1900, 1846, 1757)

no_days <- c(31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31)

df_accident <- tibble(month.name, no_accidents, no_days)
colnames(df_accident) <- c("Month", "Observed", "Number of days/Month")

df_accident["daily_ave"] <- df_accident[, 2] / df_accident[, 3]

ave_day <- sum(df_accident["Observed"]) / 365

df_accident["daily_ave_delta"] <- df_accident[, 4] - ave_day

df_accident["Expected"] <- round(ave_day * df_accident[, 3])

df_accident["Observed - Expected"] <-
  df_accident["Observed"] - df_accident["Expected"]

df_accident["Probs"] <-
  df_accident["Observed - Expected"]^2 / sum(df_accident["Observed"])

df_accident
```

```
## # A tibble: 12 x 8
##   Month      Observed 'Number of days/Month' daily_ave daily_ave_delta Expected
##   <chr>         <dbl>             <dbl>      <dbl>         <dbl>      <dbl>
## 1 January      1735                31      56.0         -4.10      1862
## 2 February    1657                28      59.2         -0.893     1682
## 3 March       1812                31      58.5         -1.62      1862
## 4 April       1962                30      65.4          5.33      1802
## 5 May         1965                31      63.4          3.32      1862
## 6 June        1849                30      61.6          1.56      1802
## 7 July        1755                31      56.6         -3.46      1862
## 8 August      1892                31      61.0          0.961     1862
## 9 September   1796                30      59.9         -0.205     1802
## 10 October    1900                31      61.3          1.22      1862
## 11 November   1846                30      61.5          1.46      1802
## 12 December   1757                31      56.7         -3.39      1862
## # ... with 2 more variables: Observed - Expected <dbl>, Probs <dbl>
```

```
X_2 = sum(df_accident["Probs"])
print(X_2)
```

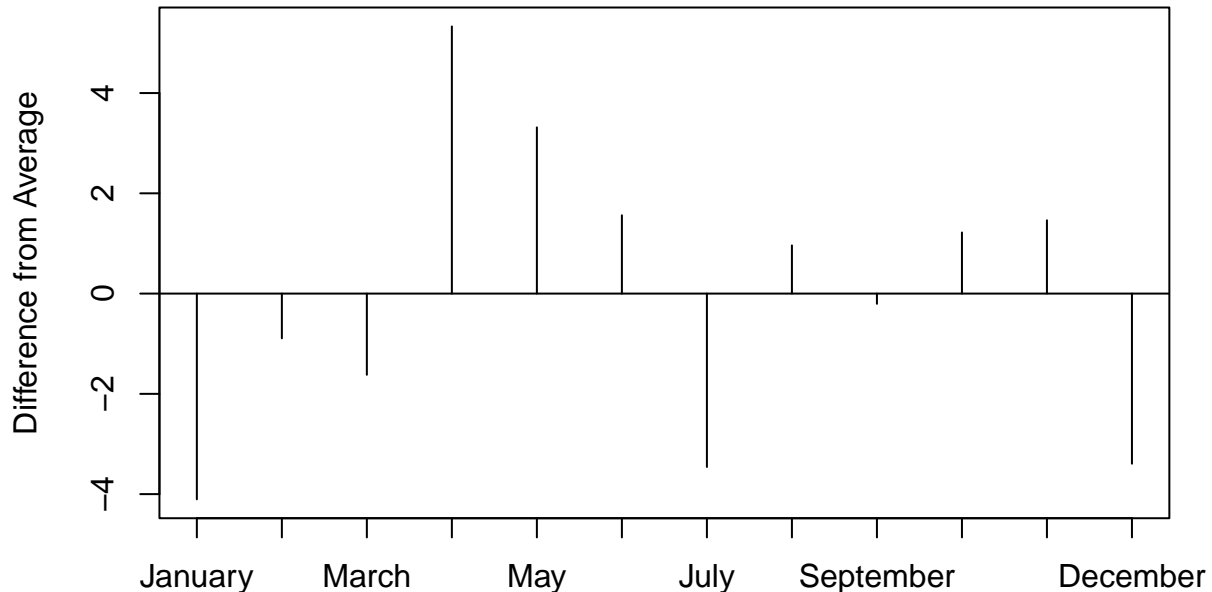
```
## [1] 3.852139
```

```
print(qt(0.95, 12 - 1 - 1))
```

```
## [1] 1.812461
```

We reject the null hypothesis because the observed test statistic is greater than the calculated value at an alpha level of 0.05. Our observed statistic is 3.852139 which is greater than 1.8124611.

```
plot(df_accident$daily_ave_delta, type = "h", xaxt = "n", xlab = "",
     ylab = "Difference from Average")
axis(1, at = 1:12, labels = df_accident$Month)
abline(h = 0)
```



July and the winter months seem to indicate a drop in the number of accidents.

Problem 3

```
set.seed(100)

options(repr.plot.width = 5, repr.plot.height = 5, repr.plot.res = 225)

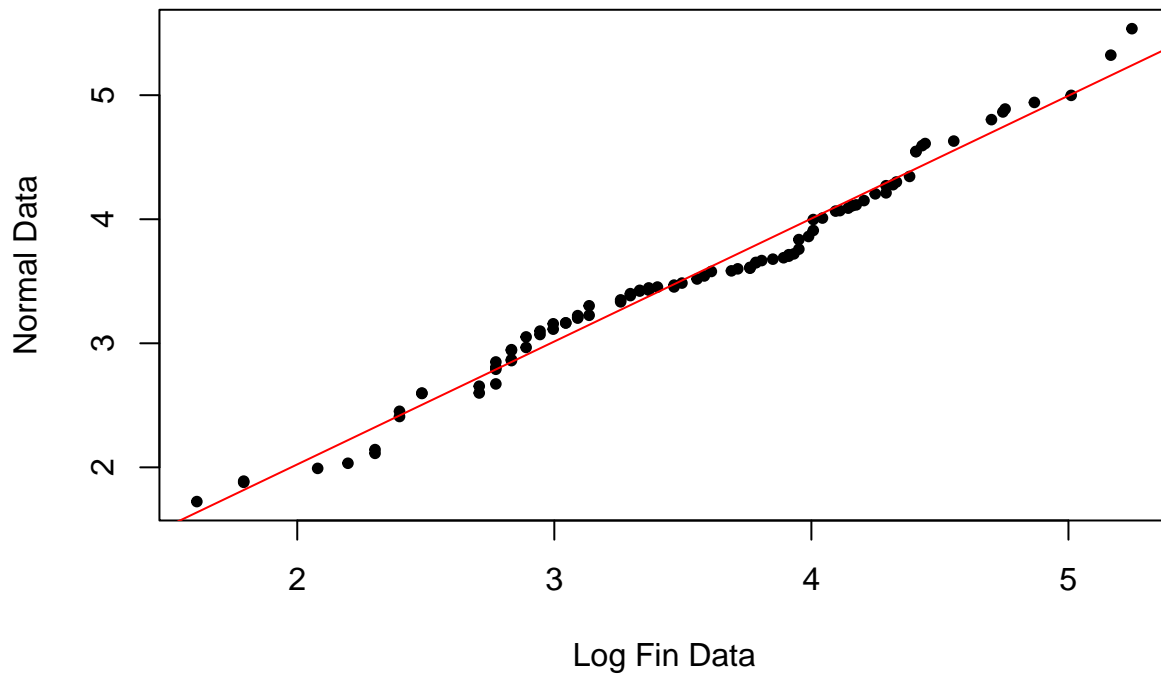
fin_data <- c(110,15,60,54,19,115,73,190,57,43,44,18,37,43,55,19,23,82,175,50,80,65,63,
             36,16,10,17,52,43,70,22,95,20,41,17,15,12,11,29,29,61,22,40,17,26,30,16,
             116,28,32,33,29,27,16,55,8,11,49,82,85,20,67,27,44,16,6,35,17,26,32,76,
             150,21,5,6,51,75,23,29,64,22,47,9,10,28,18,84,52,130,50,45,12,21,73)

fin_data <- sort(log(fin_data))

n_fin <- length(fin_data)

norm_data <- sort(rnorm(n_fin, mean = mean(fin_data), sd = sqrt(var(fin_data))))

plot(x = fin_data, y = norm_data, pch = 20, xlab = "Log Fin Data", ylab = "Normal Data")
abline(lm(norm_data ~ fin_data), col = "red")
```

Problem 4

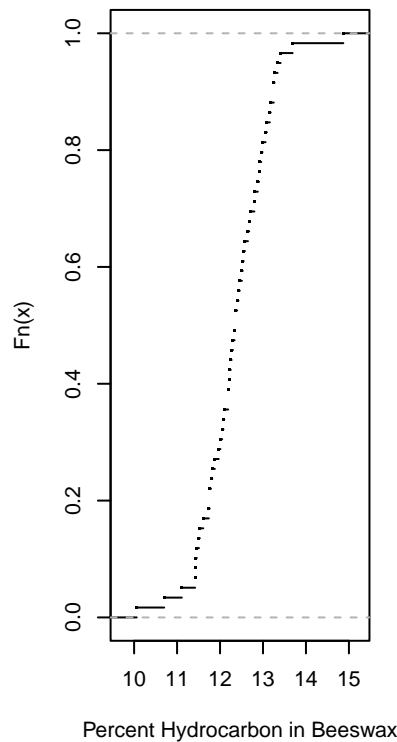
```
data <- c(12.05,12.58,10.06,14.87,12.88,10.7,13.34,13.16,12.93,11.76,12.68,
13.69,12.3,13.41,11.61,11.44,11.76,12.25,12.43,12.51,12.96,12.27,
12.34,12.81,13.18,12.46,11.11,12.19,11.97,12.99,12.53,12.19,11.82,
11.46,13.09,12.1,11.42,12.55,12.08,12.4,11.88,13.25,11.51,11.43,
12.8,11.79,12.7,13.25,11.53,12.65,13.06,12.21,11.74,12.35,13.27,
12.91,12.01,12.22,12.35)

# plots
options(repr.plot.width = 12, repr.plot.height = 5, repr.plot.res = 225)

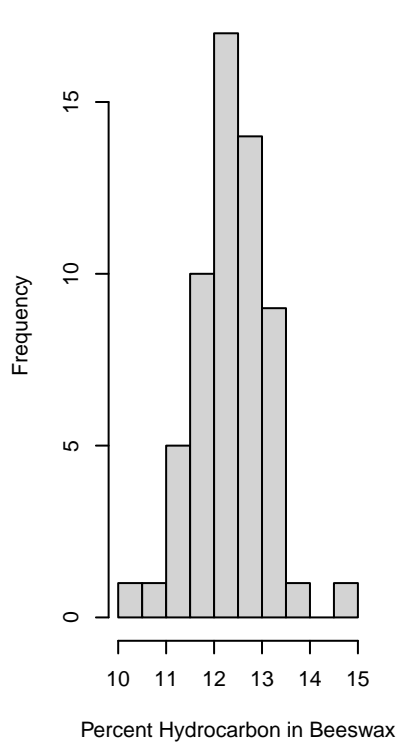
par(mfrow = c(1, 3))

ecdf_plot <- plot(ecdf(data), xlab = "Percent Hydrocarbon in Beeswax",
  main = "Beeswax Hydrocarbon Composition", pch = ".")
hist_plot <- hist(data, breaks = 12, main = "Beeswax Hydrocarbon Composition",
  xlab = "Percent Hydrocarbon in Beeswax")
qq_plot <- qqnorm(data, pch = 20, main = "Beeswax Normal QQ Plot")
qqline(data, col = "red")
```

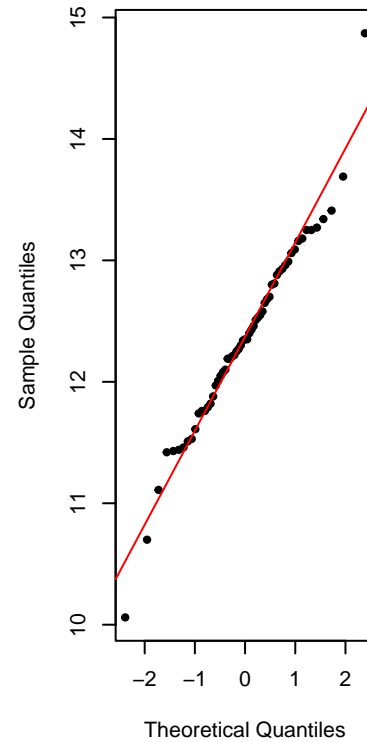
Beeswax Hydrocarbon Composition



Beeswax Hydrocarbon Composition



Beeswax Normal QQ Plot



```
# Find Quantiles
```

```
print(quantile(data, probs = c(0.9, 0.75, 0.5, 0.25, 0.1)))
```

```
##      90%      75%      50%      25%      10%
## 13.250 12.895 12.350 11.850 11.456
```

Problem 5

```
df_data <- read_csv("oldfaithful.csv")
```

```
##
## -- Column specification -----
## cols(
##   DAY = col_double(),
##   INTERVAL = col_double(),
##   DURATION = col_double()
## )
```

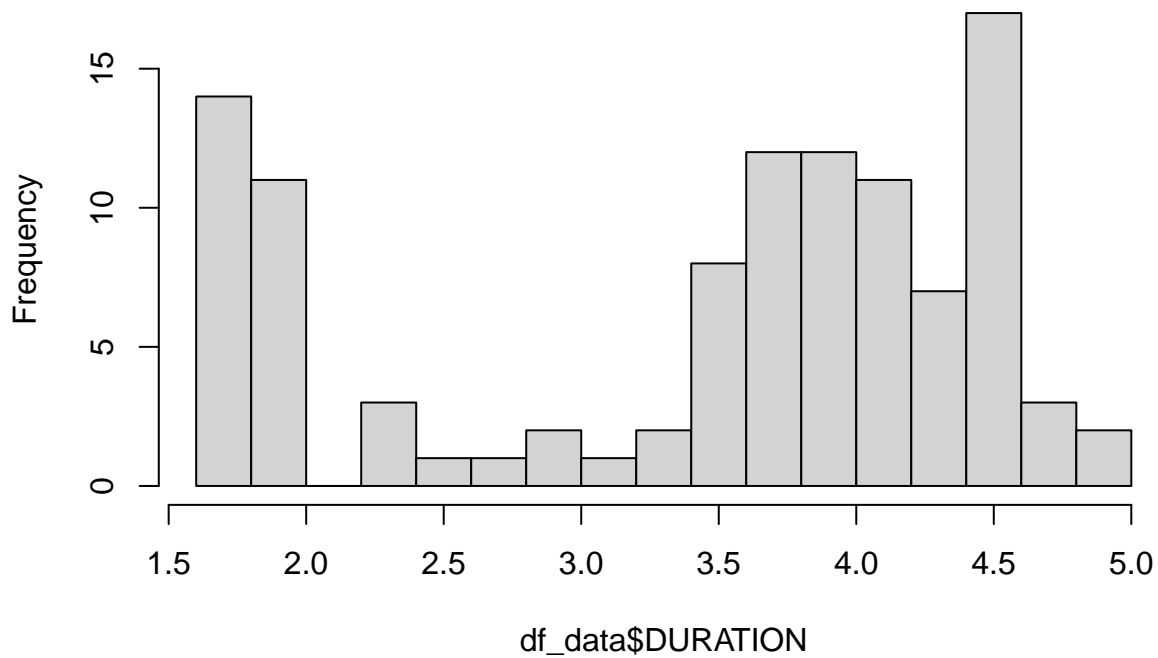
```
str(df_data)
```

```
## spec_tbl_df [107 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ DAY      : num [1:107] 1 1 1 1 1 1 1 1 1 1 ...
## $ INTERVAL: num [1:107] 78 74 68 76 80 84 50 93 55 76 ...
```

```
## $ DURATION: num [1:107] 4.4 3.9 4 4 3.5 ...
## - attr(*, "spec")=
## .. cols(
## ..   DAY = col_double(),
## ..   INTERVAL = col_double(),
## ..   DURATION = col_double()
## .. )
```

```
hist(df_data$DURATION, breaks = 12)
```

Histogram of df_data\$DURATION



```
hist(df_data$INTERVAL, breaks = 15)
```

Histogram of df_data\$INTERVAL

