Homework 3

David Hinds

2023-10-18

Part A

```
# Read in the data
url <- "https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/ThicknessGauge.dat"
thickness_gauge <- read.table(url, header = F, skip = 2, fill = T, stringsAsFactors = F)
# Note: the first row of ThicknessGauge.dat was removed manually.
# For each operator, two columns correspond to the measurements taken by that operator. This column nam
names(thickness_gauge) <- c("Part", "1_1", "1_2", "2_1", "2_2", "3_1", "3_2")
# Reshape the data
reshaped_thickness_gauge <- pivot_longer(thickness_gauge, cols = 2:7, names_to = "measurement", values_
# From the value in each measurement column, can get the Operator used to take the measurement, as well
reshaped_thickness_gauge$Operator <- as.numeric(substring(reshaped_thickness_gauge$measurement, 1, 1))
reshaped_thickness_gauge$trial <- as.numeric(substring(reshaped_thickness_gauge$measurement, 3, 3))
# Get rid of measurement column as it is no longer necessary
reshaped_thickness_gauge <- reshaped_thickness_gauge %>%
  select(Part, Operator, trial, length)

reshaped_thickness_gauge
```

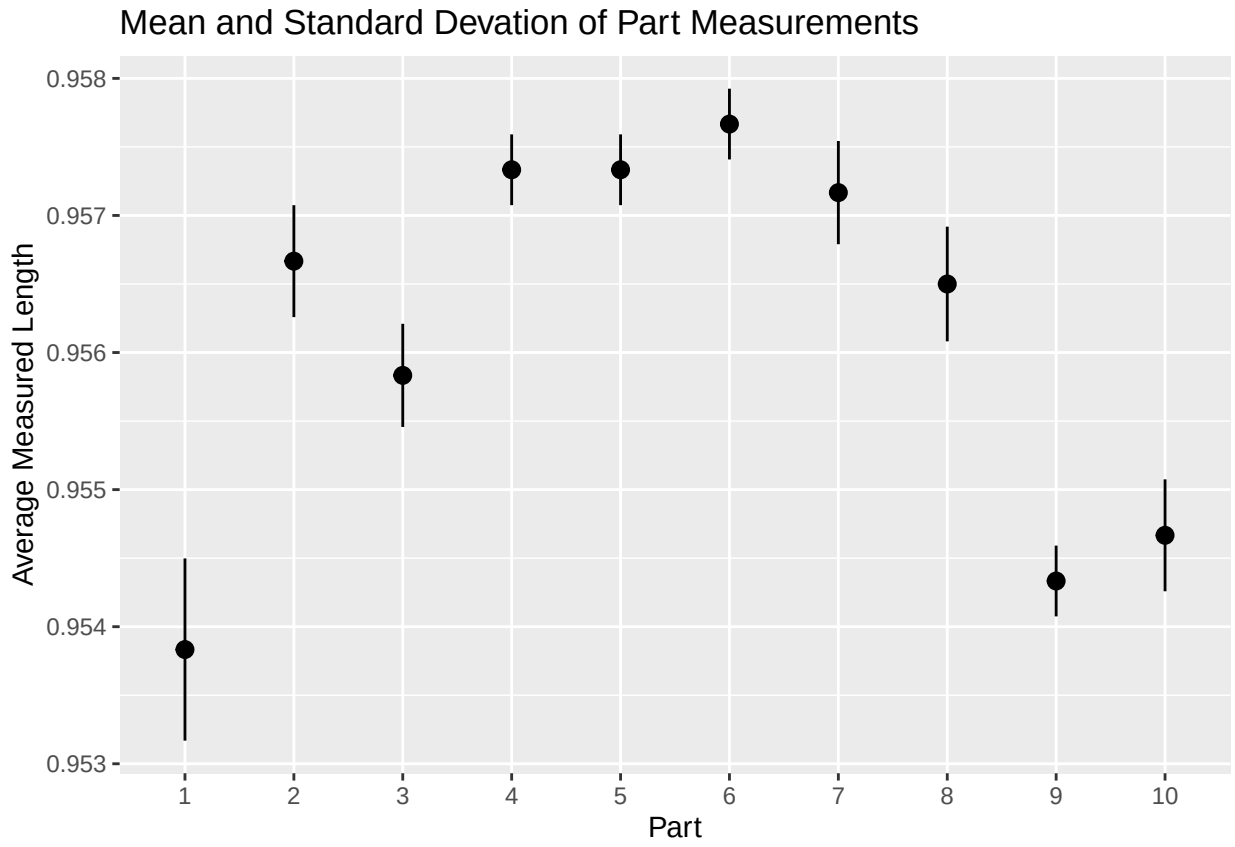
```
## # A tibble: 60 x 4
##     Part Operator trial length
##    <int>    <dbl> <dbl>  <dbl>
##  1     1        1     1  0.953
##  2     1        1     2  0.952
##  3     1        2     1  0.954
##  4     1        2     2  0.954
##  5     1        3     1  0.954
##  6     1        3     2  0.956
##  7     2        1     1  0.956
##  8     2        1     2  0.956
##  9     2        2     1  0.956
## 10     2        2     2  0.957
## # i 50 more rows
```

```
# Data summary: for each part, would like the mean of all of the measurements taken from it.
thickness_summary <- reshaped_thickness_gauge %>%
  group_by(Part) %>%
  summarise(avg_meas_length = mean(length), meas_sd = sd(length))

thickness_summary
```

```
## # A tibble: 10 x 3
##     Part avg_meas_length  meas_sd
```

```
##    <int>           <dbl>    <dbl>
##  1     1           0.954 0.00133
##  2     2           0.957 0.000816
##  3     3           0.956 0.000753
##  4     4           0.957 0.000516
##  5     5           0.957 0.000516
##  6     6           0.958 0.000516
##  7     7           0.957 0.000753
##  8     8           0.956 0.000837
##  9     9           0.954 0.000516
## 10    10           0.955 0.000816
```

```
thickness_summary$thickness_LB = thickness_summary$avg_meas_length - thickness_summary$meas_sd/2

thickness_summary$thickness_UB = thickness_summary$avg_meas_length + thickness_summary$meas_sd/2

ggplot(data = thickness_summary) + geom_pointrange(aes(x = as.factor(Part), y = avg_meas_length, ymin =
  labs(title = "Mean and Standard Devation of Part Measurements", x = "Part", y = "Average Measured Leng
```



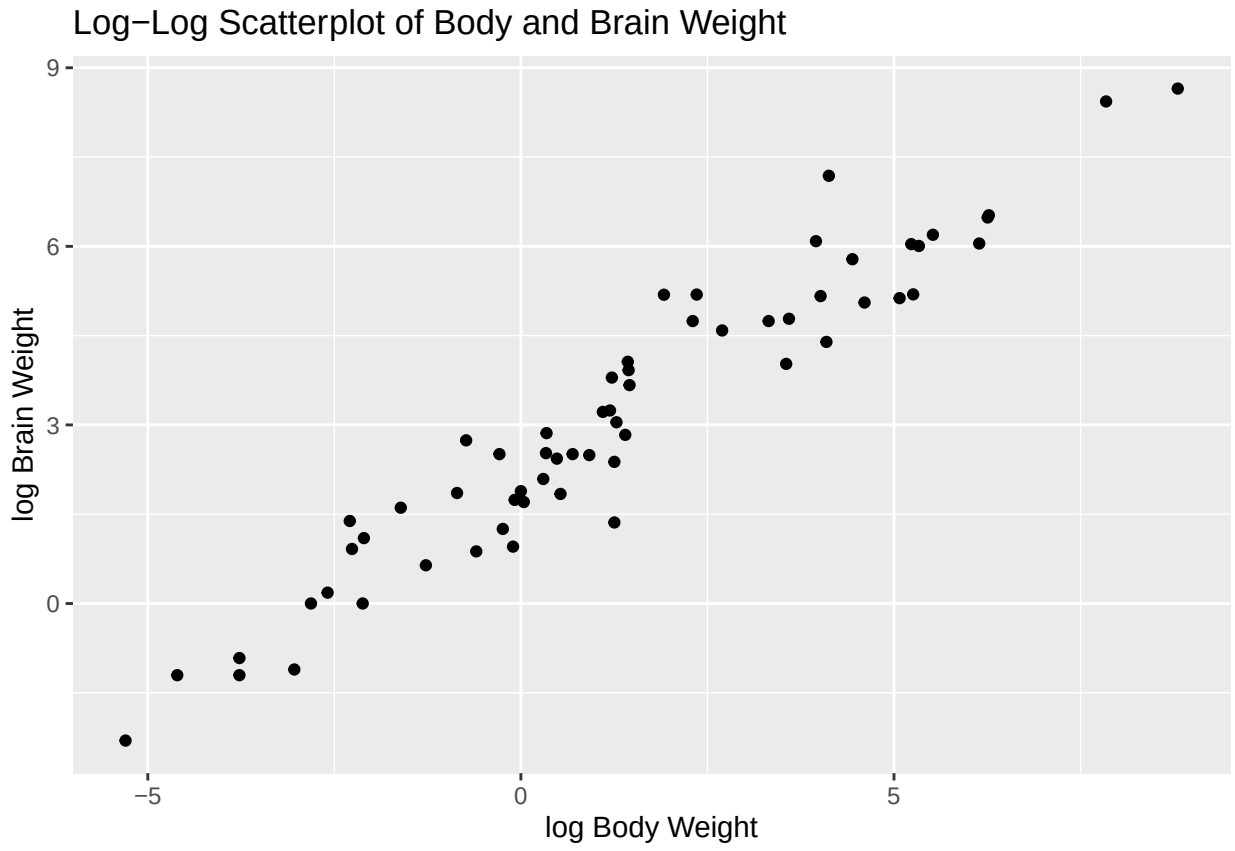
Part B

```
# Read in the data
url <- "https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BrainandBodyWeight.dat"
brain_body_wt <- read.table(url, header = F, skip = 1, fill = T, stringsAsFactors = F)
# Name the columns appropriately.
names(brain_body_wt) <- c("BodyWt1", "BrainWt1", "BodyWt2", "BrainWt2", "BodyWt3", "BrainWt3")
```

```r
# Reshape the data
body_wt <- pivot_longer(brain_body_wt, cols = c(1,3,5), values_to = "BodyWt")
brain_wt <- pivot_longer(brain_body_wt, cols = c(2,4,6), values_to = "BrainWt")
reshaped_brain_body_wt <- data.frame(BodyWt = body_wt$BodyWt, BrainWt = brain_wt$BrainWt) %>%
  filter(!is.na(BodyWt) & !is.na(BrainWt))
# Summary table: mean and sd of the brain and body weights of the subjects.
reshaped_brain_body_wt %>%
  summarise(body_wt_mean = mean(BodyWt), body_wt_sd = sd(BodyWt),
            brain_wt_mean = mean(BrainWt), brain_wt_sd = sd(BrainWt))
```

```
##   body_wt_mean body_wt_sd brain_wt_mean brain_wt_sd
## 1       198.79    899.158      283.1344    930.2789
```

```r
ggplot(data = reshaped_brain_body_wt) +
  geom_point(aes(x = log(BodyWt), y = log(BrainWt))) +
  labs(title = "Log-Log Scatterplot of Body and Brain Weight", x = "log Body Weight", y = "log Brain Wei
```



Part C

```r
# Read in the data
url = "https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LongJumpData.dat"
long_jump <- fread(url, fill = TRUE, skip = 1)
# Reshape the data
years <- long_jump[,c(1,3,5,7)]
distances <- long_jump[,c(2,4,6,8)]
years_long <- pivot_longer(years, cols = 1:4, values_to = "Year")
```

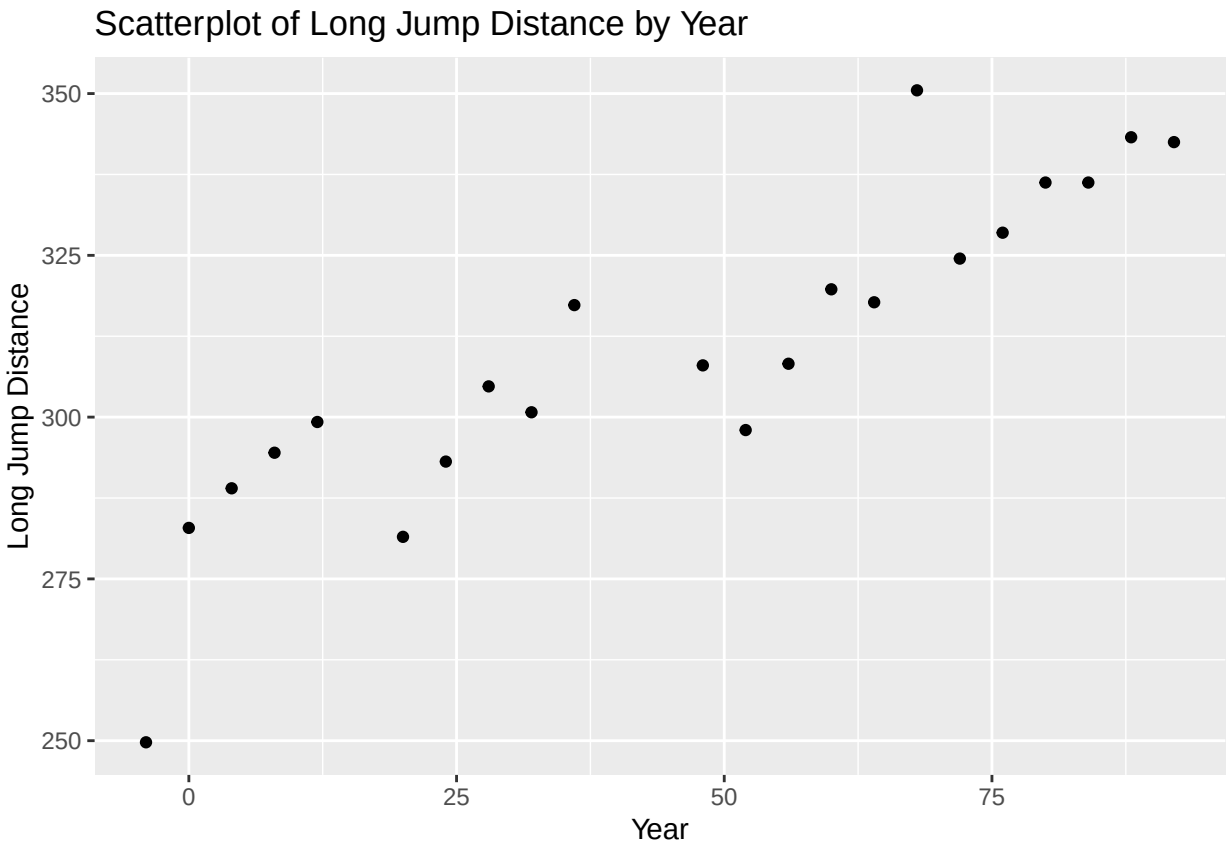
```r
distances_long <- pivot_longer(distances, cols = 1:4, values_to = "Distance")
reshaped_long_jump <- data.frame(Year = years_long$Year, Distance = distances_long$Distance) %>%
  filter(!is.na(Year) & !is.na(Distance))
reshaped_long_jump <- reshaped_long_jump[order(reshaped_long_jump$Year),]

reshaped_long_jump
```

```
##    Year Distance
## 1    -4   249.75
## 5     0   282.88
## 9     4   289.00
## 13    8   294.50
## 17   12   299.25
## 20   20   281.50
## 2    24   293.13
## 6    28   304.75
## 10   32   300.75
## 14   36   317.31
## 18   48   308.00
## 21   52   298.00
## 3    56   308.25
## 7    60   319.75
## 11   64   317.75
## 15   68   350.50
## 19   72   324.50
## 22   76   328.50
## 4    80   336.25
## 8    84   336.25
## 12   88   343.25
## 16   92   342.50
```

```r
# Summary table: the mean and sd of the long jump distances
reshaped_long_jump %>%
  summarise(distance_mean = mean(Distance), distance_sd = sd(Distance))
```

```
##   distance_mean distance_sd
## 1      310.2873    24.36121
```

```r
ggplot(data = reshaped_long_jump) +
  geom_point(aes(x = Year, y = Distance)) +
  labs(title = "Scatterplot of Long Jump Distance by Year", x = "Year", y = "Long Jump Distance")
```

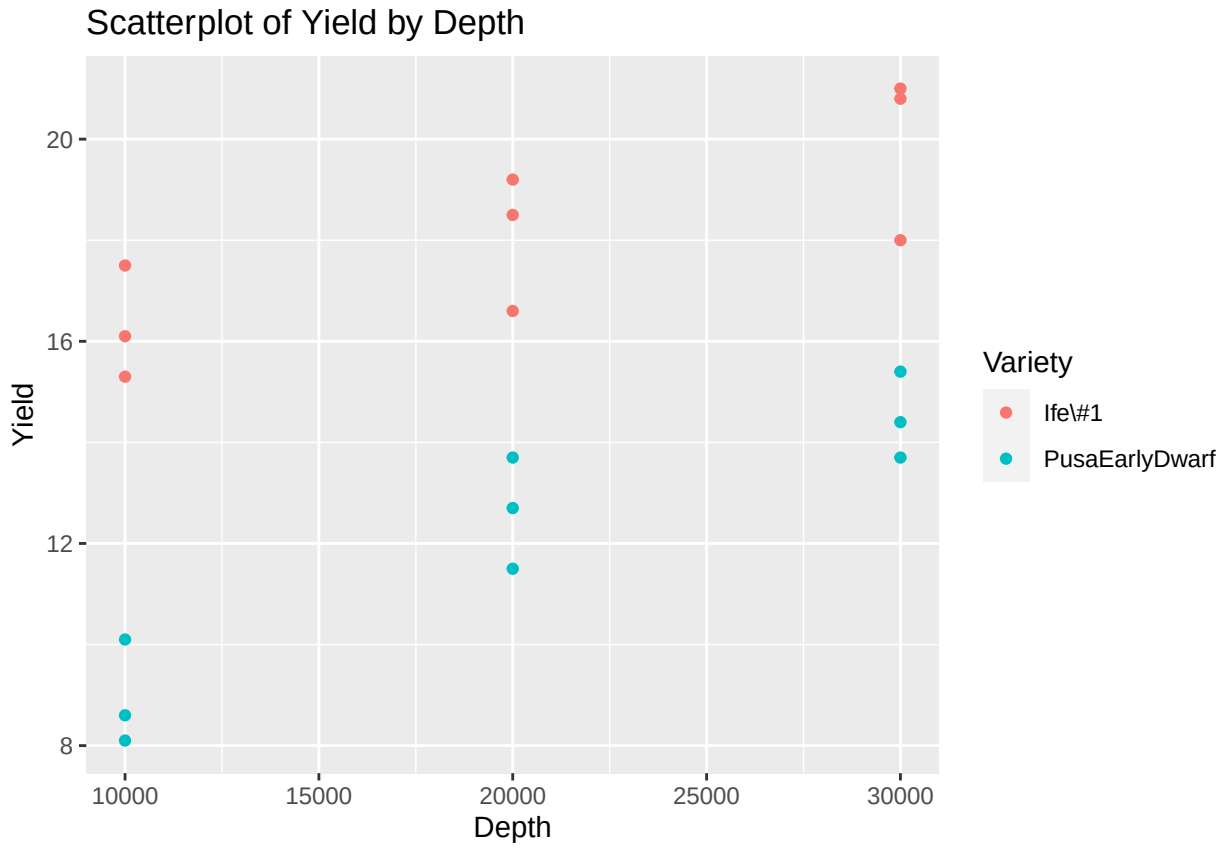


Part D

```
# Read in the data
url = "https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/tomato.dat"
lines <- gsub(",", " ", readLines(url)[-2:-1])
text = lines[1]
for(i in 2:length(lines)){
  text <- paste(text, lines[i], sep = "\n")
}
tomato <- fread(text)
# Reshape the data
names(tomato) <- c("Variety", "1_1", "1_2", "1_3",
                              "2_1", "2_2", "2_3",
                              "3_1", "3_2", "3_3")
reshaped_tomato <- pivot_longer(tomato, cols = 2:10, names_to = "measurement", values_to = "Yield") %>%
  mutate(Depth = as.numeric(substring(measurement, 1, 1))*10000,
         trial = as.numeric(substring(measurement, 3, 3))) %>%
  select(Variety, Depth, trial, Yield)

reshaped_tomato

# Summary Table: the mean yield of each variety of tomato at each depth
reshaped_tomato %>%
  group_by(Variety, Depth) %>%
  summarise(mean_meas_yield = mean(Yield))
```

```
## `summarise()` has grouped output by 'Variety'. You can override using the
## `.groups` argument.
```

```
# Unfortunately summarise gives unwanted console output here.

ggplot(data = reshaped_tomato) +
  geom_point(aes(x = Depth, y = Yield, color = Variety)) +
  labs(title = "Scatterplot of Yield by Depth", x = "Depth", y = "Yield")
```



Part E

```
# Read in the data
url = "https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LarvaeControl.dat"
larvae <- fread(url, fill = TRUE, skip = 2)
# Reshape the data
names(larvae)[2:11] <- c("1_1", "1_2", "1_3","1_4","1_5",
                         "2_1", "2_2", "2_3","2_4","2_5")
reshaped_larvae <- pivot_longer(larvae, cols = 2:11, names_to = "measurement", values_to = "larvae") %>%
  mutate(Age = as.numeric(substring(measurement, 1, 1)),
         Treatment = as.numeric(substring(measurement, 3, 3))) %>%
  select(Block, Age, Treatment, larvae)  # Get rid of measurement column

reshaped_larvae

# Summary Table: mean larvae count for each age and treatment
reshaped_larvae %>%
```

```
  group_by(Age, Treatment) %>%
  summarise(mean_larvae = mean(larvae))
```

```
## `summarise()` has grouped output by 'Age'. You can override using the `.groups`
## argument.
```

```
ggplot(data = reshaped_larvae) +
  geom_boxplot(aes(x = as.factor(Treatment), y = larvae, color = as.factor(Age))) +
  labs(title = "Boxplots of Larvae Count by Treatment", x = "Treatment", y = "Larvae Count", color = "Ag
```