HW@

Chris Conte

2023-10-18

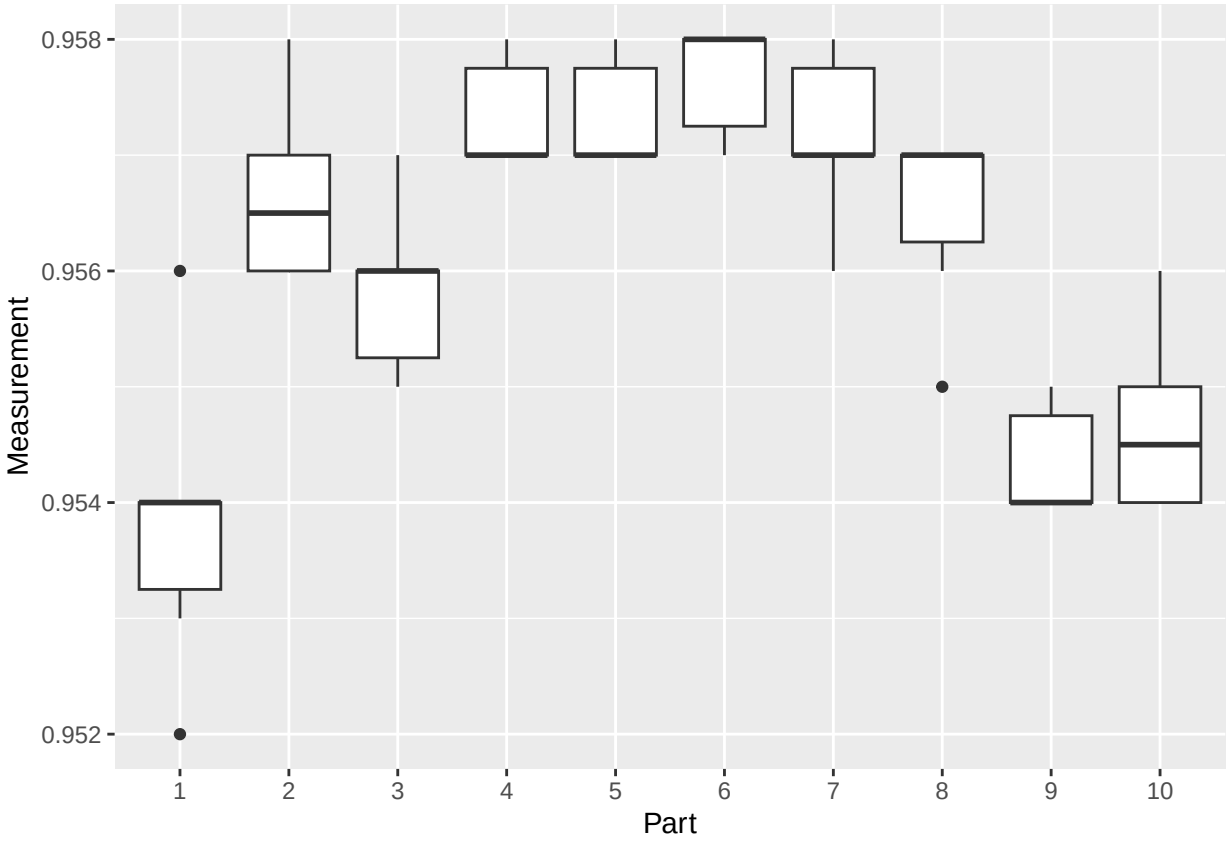
```
library(dplyr)
library(tidyr)
library(ggplot2)
library(data.table)
```

Part A

```
raw_data<-read.table("https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/ThicknessGauge.dat", heade
cleaned_data = raw_data %>% pivot_longer(V1, values_to='part', names_to=NULL) %>% pivot_longer(-part, na
cleaned_data %>% group_by(part) %>% summarise(avg_measurement = mean(measurement))
```

```
## # A tibble: 10 x 2
##     part avg_measurement
##    <int>           <dbl>
##  1     1           0.954
##  2     2           0.957
##  3     3           0.956
##  4     4           0.957
##  5     5           0.957
##  6     6           0.958
##  7     7           0.957
##  8     8           0.956
##  9     9           0.954
## 10    10           0.955
```

```
ggplot(data=cleaned_data, aes(x=as.factor(part), y= measurement))+geom_boxplot()+labs(x='Part', y='Meas
```

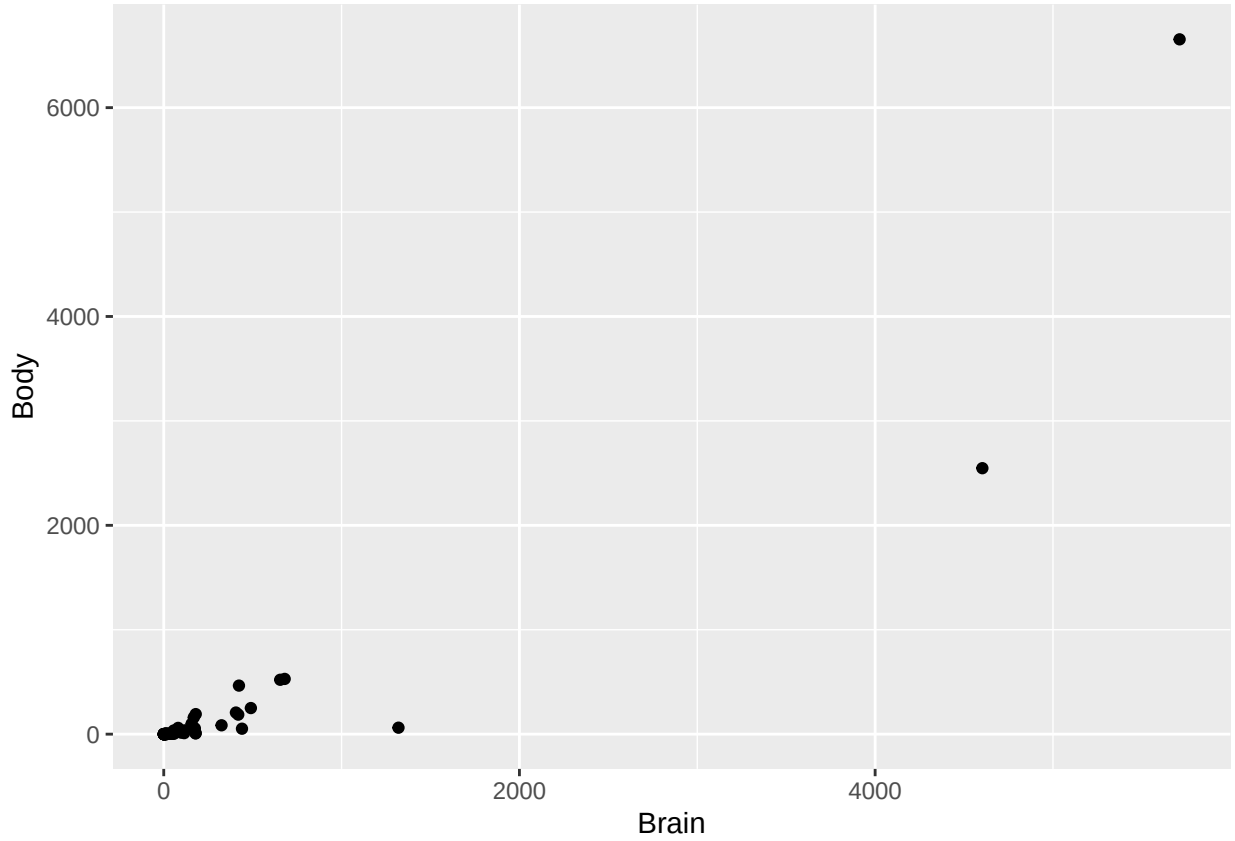


Part B

```
raw_data<-fread('https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BrainandBodyWeight.dat', fill=TR
cleaned_data = raw_data %>% pivot_longer(c(V1,V3,V5), values_to = "Body") %>% select(Body) %>% cbind(ra
summary(cleaned_data)
```

```
##       Body              Brain        
##  Min.   :   0.005   Min.   :   0.10  
##  1st Qu.:   0.600   1st Qu.:   4.25  
##  Median :   3.342   Median :  17.25  
##  Mean   : 198.790   Mean   : 283.13  
##  3rd Qu.:  48.202   3rd Qu.: 166.00  
##  Max.   :6654.000   Max.   :5712.00
```

```
ggplot(data = cleaned_data, aes(Brain, Body))+geom_point()
```



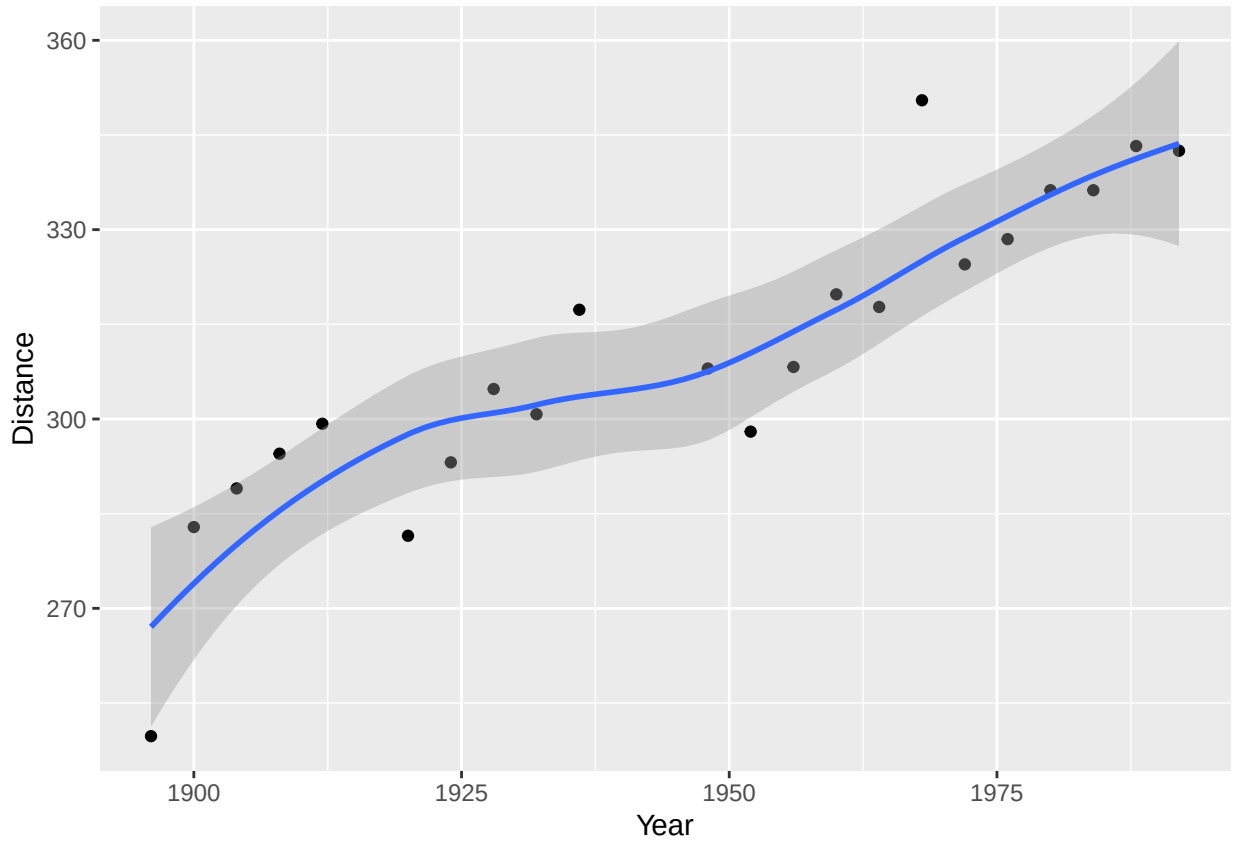
Part C

```
raw_data<-fread('https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LongJumpData.dat', fill=TRUE, he
cleaned_data = raw_data %>% pivot_longer(c(V1,V3,V5, V7), values_to = "Year") %>% mutate(Year=Year+1900)
summary(cleaned_data)
```

```
##       Year         Distance
##  Min.   :1896   Min.   :249.8
##  1st Qu.:1921   1st Qu.:295.4
##  Median :1950   Median :308.1
##  Mean   :1945   Mean   :310.3
##  3rd Qu.:1971   3rd Qu.:327.5
##  Max.   :1992   Max.   :350.5
```

```
ggplot(data = cleaned_data, aes(Year, Distance))+geom_point()+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Part D

```
raw_data<-fread('https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/tomato.dat', fill=TRUE, header=
cleaned_data = raw_data %>% filter(row_number() != 1) %>% pivot_longer(V1, values_to = "variety", names_
cleaned_data %>% group_by(variety) %>% summarise(avg_yield = mean(yield))
```

```
## # A tibble: 2 x 2
##   variety        avg_yield
##   <chr>              <dbl>
## 1 Ife#1               18.1
## 2 PusaEarlyDwarf      12.0
```

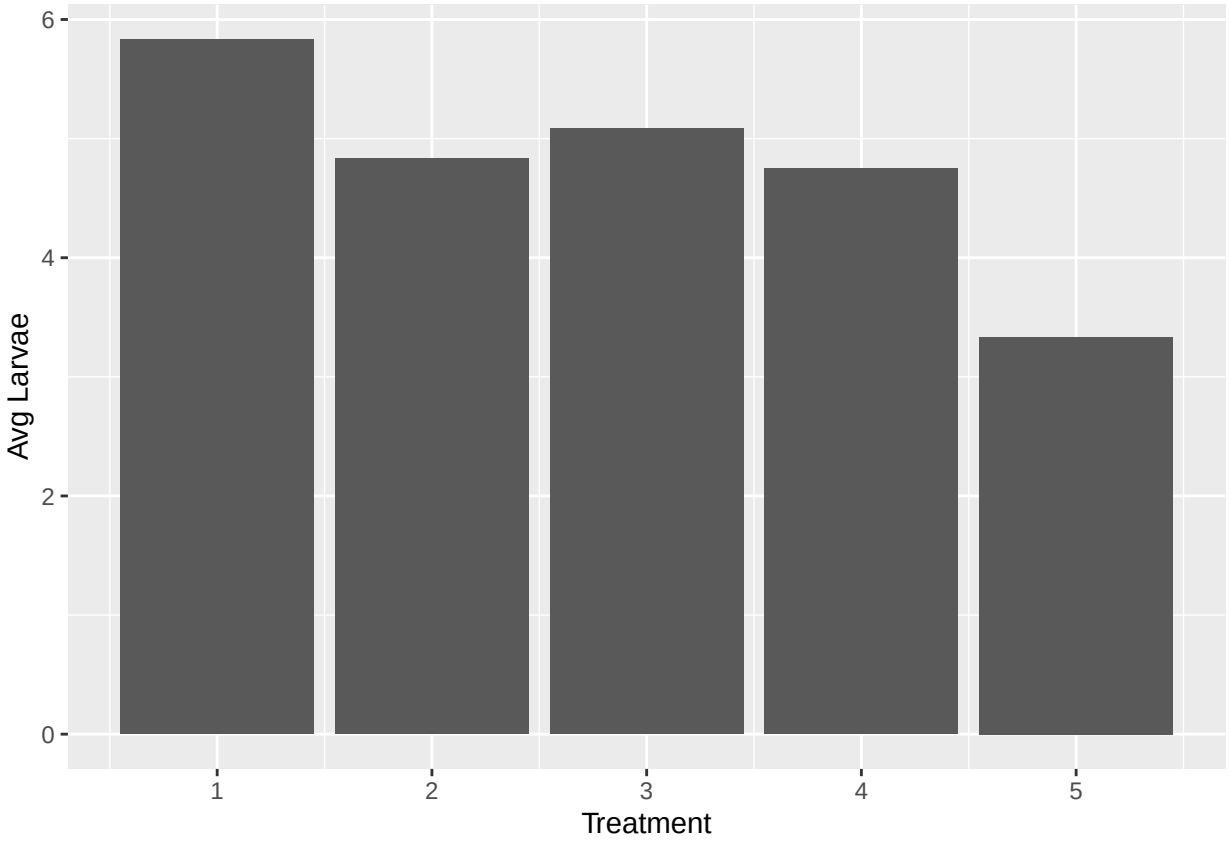
```
ggplot(data = cleaned_data, aes(x=variety, y=yield))+geom_violin()
```

Part E

```
raw_data<-fread('https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LarvaeControl.dat', fill=TRUE, l
cleaned_data = raw_data %>% pivot_longer(V1, values_to='block', names_to = NULL) %>% pivot_longer(-block
summary_data = cleaned_data %>% filter(block>2) %>% group_by(treatment) %>% summarise(avg_larvae = mean
summary_data
```

```
## # A tibble: 5 x 2
##   treatment avg_larvae
##       <dbl>      <dbl>
## 1         1       5.83
## 2         2       4.83
## 3         3       5.08
## 4         4       4.75
## 5         5       3.33
```

```
ggplot(data = summary_data, aes(x=treatment, y = avg_larvae))+geom_bar(stat='identity')+labs(x='Treatmen
```


6
4
2
0
Avg Larvae
1
2
3
4
5
Treatment