# Sarah_Widener_W7.Rmd

Sarah_Widener_W7.Rmd

10/12/2020

#Using the pheno.txt inside the HW4 folder, do the following:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v stringr 1.4.0
## v tidyr   1.1.2     v forcats 0.5.0
## v readr   1.3.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
HW7data <- read.table("https://raw.githubusercontent.com/IntroToR/class_notes/master/HW4/pheno.txt", hea
head(HW7data)
```

```
##    LOC    GENO     TAXA range row BLOCK         Y         M    h1   h2   h3
## 1 12EF    Mila PI506058     1   1     1 13.245637 0.6210611 0.210 1.25 2.30
## 2 12EF PRE0465 PI155138     1   2     1 12.253702 0.6716969 0.264 1.25 2.20
## 3 12EF PRE1022 PI329902     1   3     1  4.241176 0.7613764 0.265 1.15 2.05
## 4 12EF PRE0139 PI276790     1   4     1  7.478768 0.7423784 0.247 1.40 2.45
## 5 12EF PRE0237 PI570719     1   5     1  4.498335 0.7550720 0.321 1.20 1.95
## 6 12EF PRE0028 PI148089     1   6     1  9.348920 0.7644761 0.298 1.30 2.55
```

```
##     h4
## 1 3.30
## 2 3.20
## 3 2.40
## 4 3.70
## 5 2.65
## 6 3.75
```

```
library(tidyr)
HW7data2 <- pivot_longer(HW7data, cols = c("h1", "h2", "h3", "h4"), names_to = "h")
head(HW7data2)
```

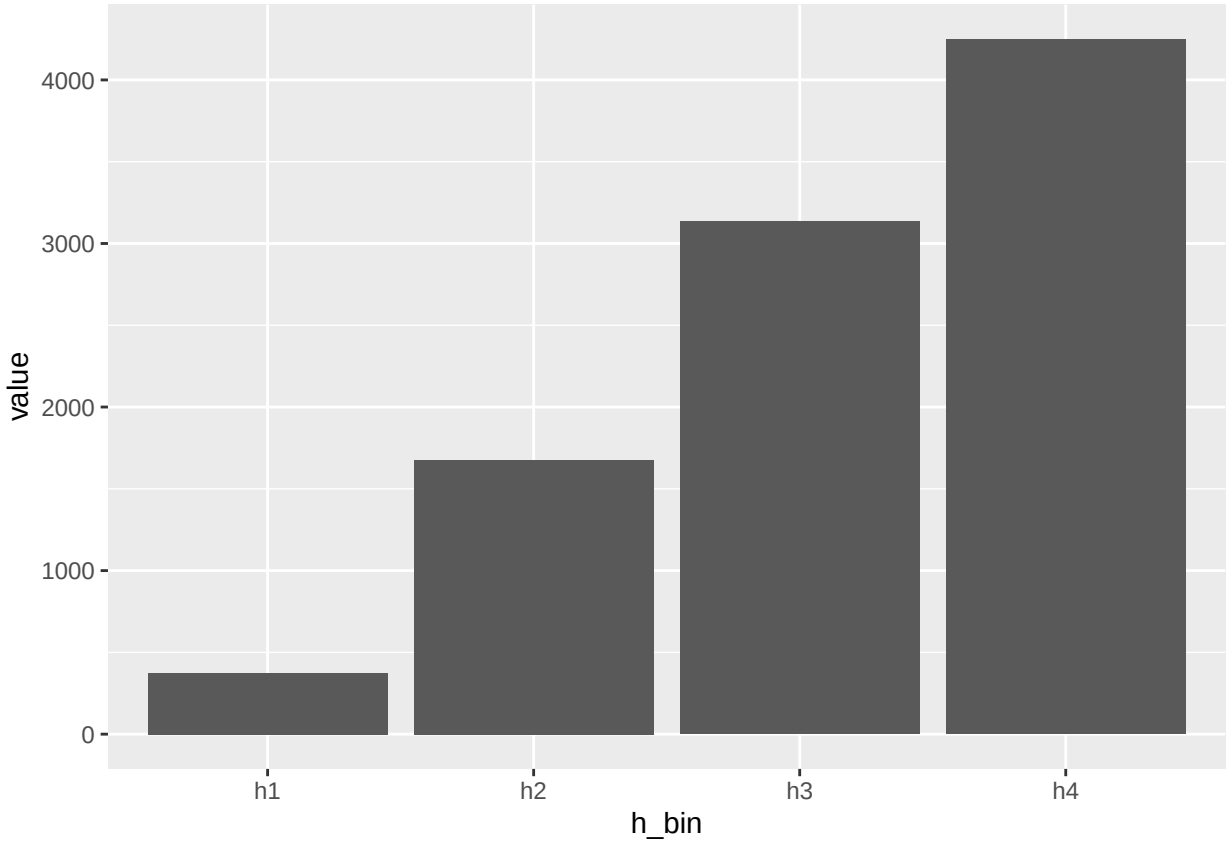
```
## # A tibble: 6 x 10
##   LOC   GENO    TAXA     range   row BLOCK     Y     M h     value
##   <fct> <fct>   <fct>    <int> <int> <int> <dbl> <dbl> <chr> <dbl>
## 1 12EF  Mila    PI506058     1     1     1  13.2 0.621 h1    0.21
## 2 12EF  Mila    PI506058     1     1     1  13.2 0.621 h2    1.25
## 3 12EF  Mila    PI506058     1     1     1  13.2 0.621 h3    2.3
## 4 12EF  Mila    PI506058     1     1     1  13.2 0.621 h4    3.3
## 5 12EF  PRE0465 PI155138     1     2     1  12.3 0.672 h1    0.264
## 6 12EF  PRE0465 PI155138     1     2     1  12.3 0.672 h2    1.25
```

#1) A histogram with h1, h2, h3, and h4. (Tip: you need to transform this data in a long format)

```
library(ggplot2)
ggplot(HW7data2, aes(x = h, y = value)) +
  geom_histogram(stat = "identity") +
  labs(x = "h_bin", y = "value")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
## Warning: Removed 153 rows containing missing values (position_stack).
```



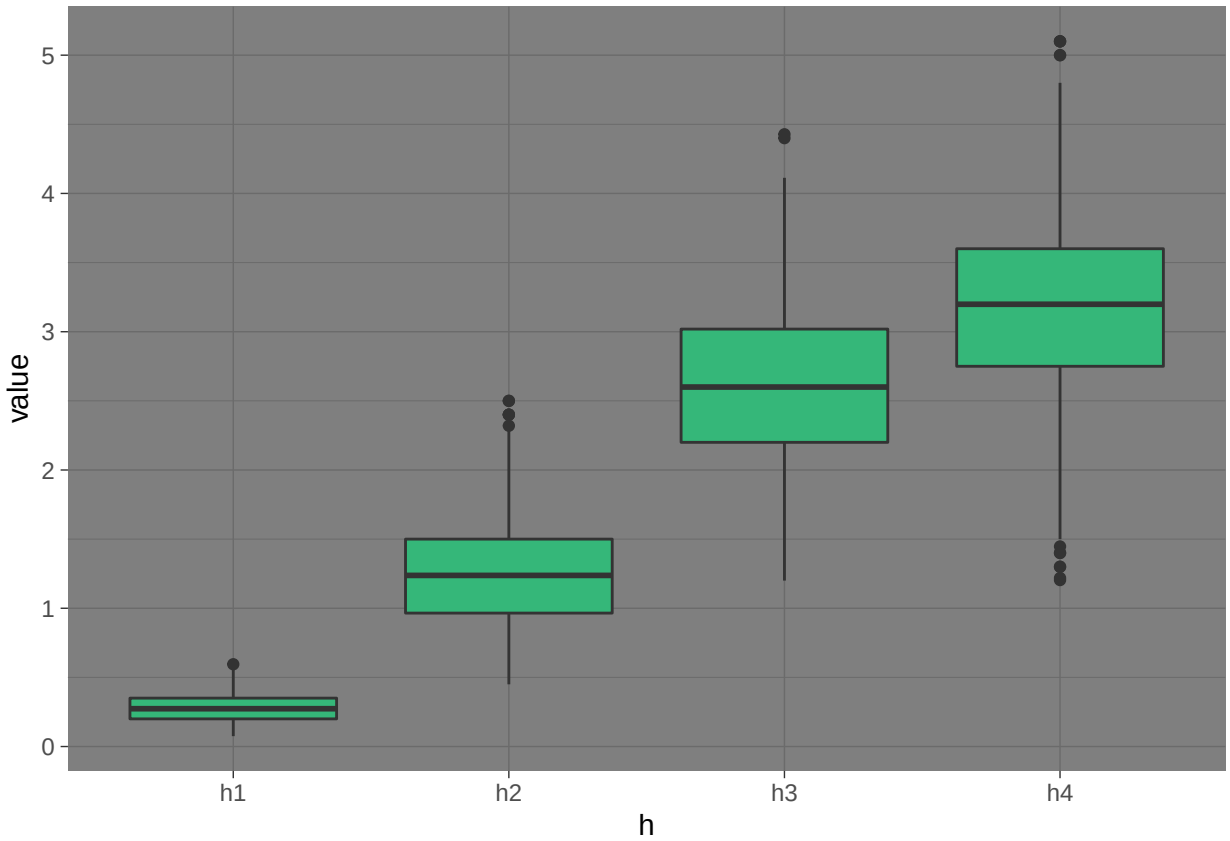

#2) Using the variable "Y" as the response variable, create a boxplot for each value of row and range. Make one plot for row and another plot for range using a facet option.

```
HW7data3 <- pivot_longer(HW7data, row:range,names_to = "type", values_to = "values")
head(HW7data3)
```

```
## # A tibble: 6 x 12
##   LOC   GENO    TAXA     BLOCK     Y     M    h1    h2    h3    h4 type  values
##   <fct> <fct>   <fct>    <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>  <int>
## 1 12EF  Mila    PI506058     1 13.2  0.621 0.21   1.25  2.3    3.3 row        1
## 2 12EF  Mila    PI506058     1 13.2  0.621 0.21   1.25  2.3    3.3 range      1
## 3 12EF  PRE0465 PI155138     1 12.3  0.672 0.264  1.25  2.2    3.2 row        2
## 4 12EF  PRE0465 PI155138     1 12.3  0.672 0.264  1.25  2.2    3.2 range      1
## 5 12EF  PRE1022 PI329902     1  4.24 0.761 0.265  1.15  2.05   2.4 row        3
## 6 12EF  PRE1022 PI329902     1  4.24 0.761 0.265  1.15  2.05   2.4 range      1
```

```
ggplot(HW7data2) +
 aes(x = h, y = value) +
 geom_boxplot(fill = "#35b779") +
 theme_dark()
```

```
## Warning: Removed 153 rows containing non-finite values (stat_boxplot).
```




#3) Calculate the mean of each TAXA, filter the TAXA with a mean greater than 12, and create a barplot with the error bar on top of it (use the value 1 as the standard error).

```
HW7data2%>%
  group_by(TAXA)%>%
  summarise_at(vars(Y),
  list(name=mean))
```

```
## # A tibble: 480 x 2
##    TAXA               name
##    <fct>             <dbl>
##  1 commercial_hybrid NA
##  2 NSL102169          6.76
##  3 NSL102183          4.51
##  4 NSL360516         11.0
##  5 NSL360526          6.11
##  6 NSL360555          7.42
##  7 NSL365683         11.7
##  8 NSL365686          6.65
##  9 NSL365693          4.62
## 10 NSL365694          4.00
## # ... with 470 more rows
```

```
mean2 <- HW7data2%>%
  group_by(TAXA)%>%
  summarise_at(vars(Y),
```

```
  list(name=mean)) %>%
  filter(name>12)
n<-13
sd<- rep(1, each=n)
mean3<- mean2%>% add_column(sd)


finalplot <- ggplot(mean3)+
  geom_bar(aes(x = TAXA, y = name), stat="identity", fill="red", alpha=0.5)+
  geom_errorbar(aes(x=TAXA, ymin=name-sd, ymax=name+sd), colour="35b779",width=0.3, size=1)+
  xlab("TAXA")+
  ylab("Mean")+
 ggtitle("Average Y value by TAXA")
plot(finalplot)
```