Guinea Pig Problem

Anush, Ashkhen, Susie

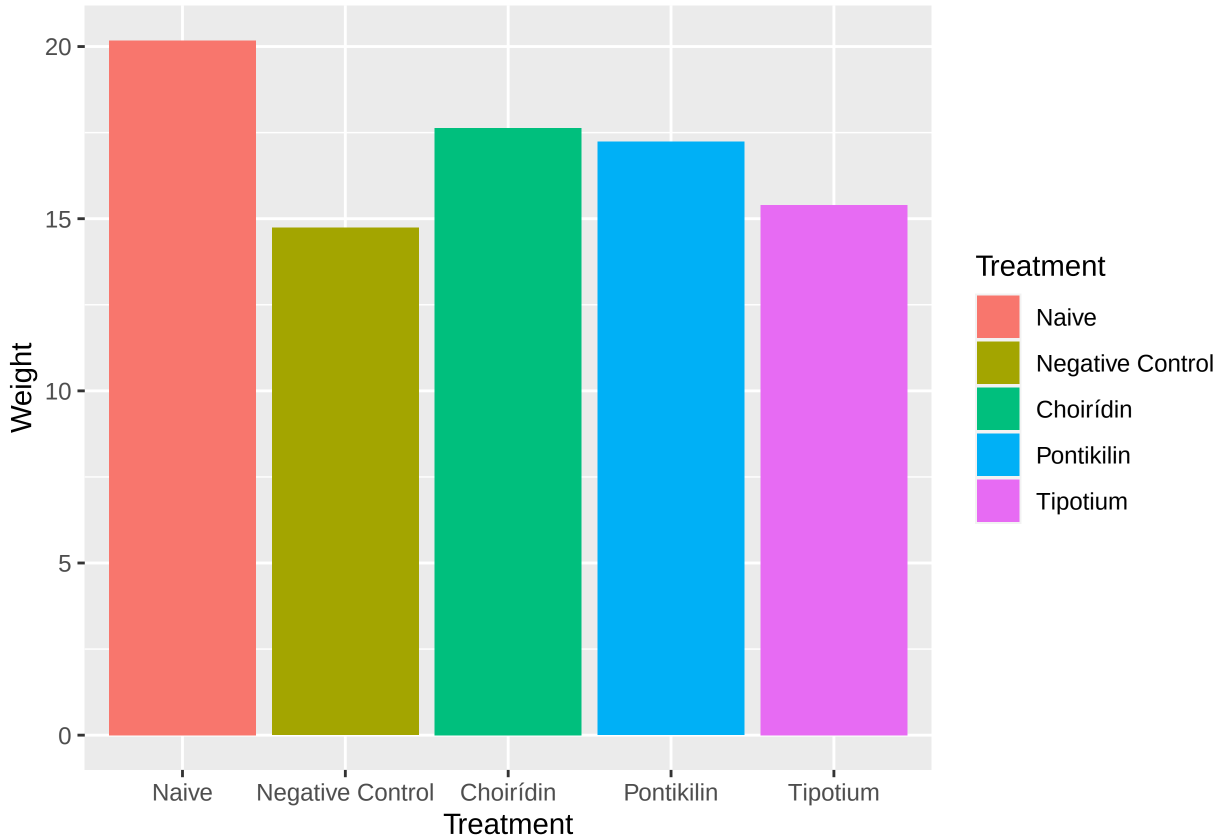
7/12/2021

- Find the data file in the "data" folder

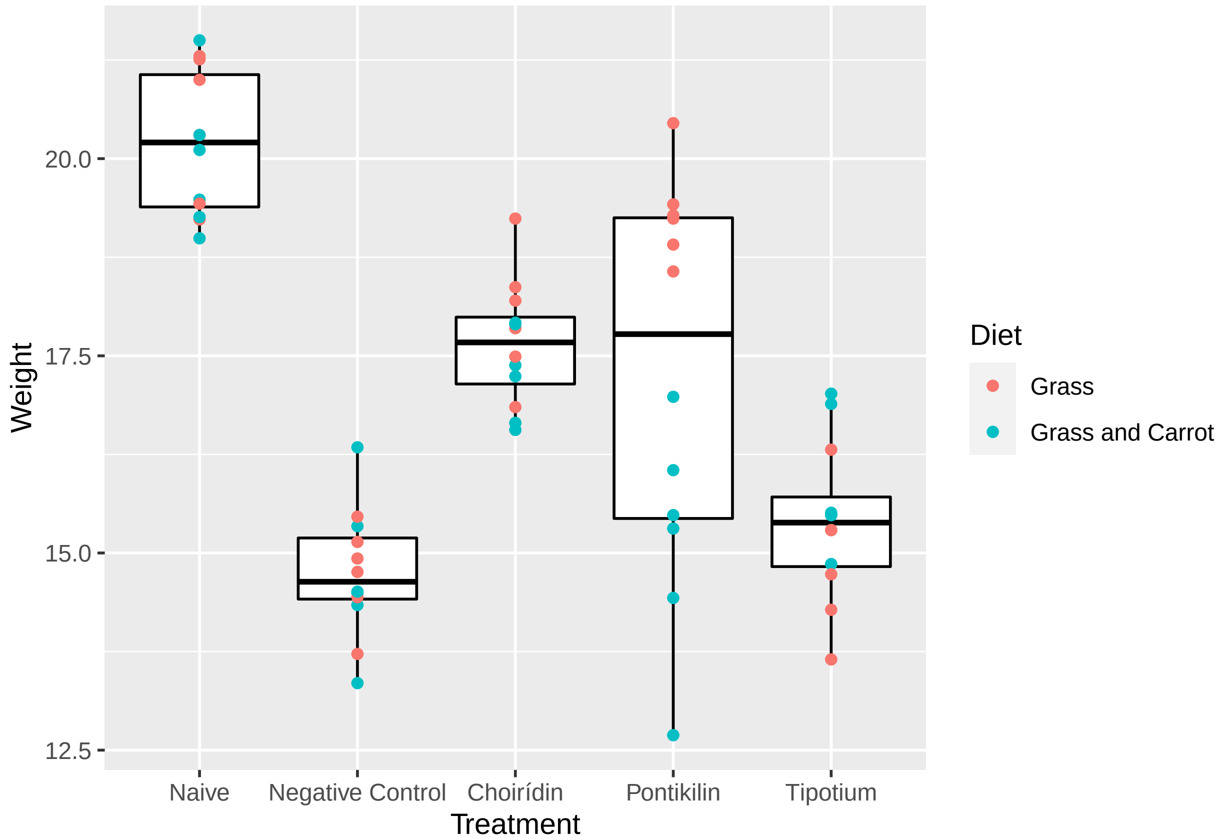
```
library(ggplot2)
library(dplyr)

data <- read.csv('dt1.csv')
```

- Visualize the data (dot plot, histogram, boxplot)

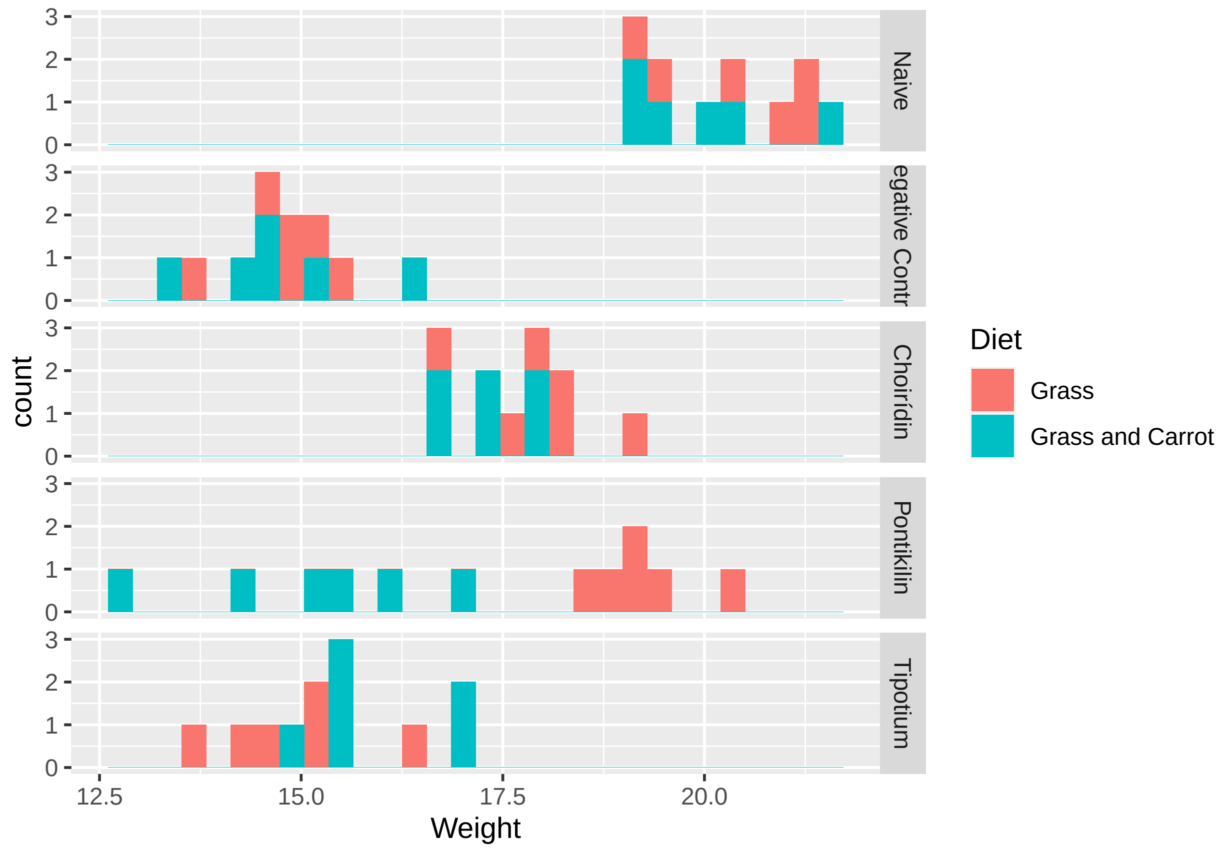
```
data$Treatment <- factor(data$Treatment,
                         levels=c("Naive", "Negative Control", "Choirídin", "Pontikilin", "Tipotium"))


dt <- group_by(data, Treatment)
d <- dt%>%summarise(Weight=mean(Weight))
p2 <- ggplot(d, aes(x=Treatment, y = Weight, fill = Treatment)) +
  geom_bar(stat = "identity")
p2
```



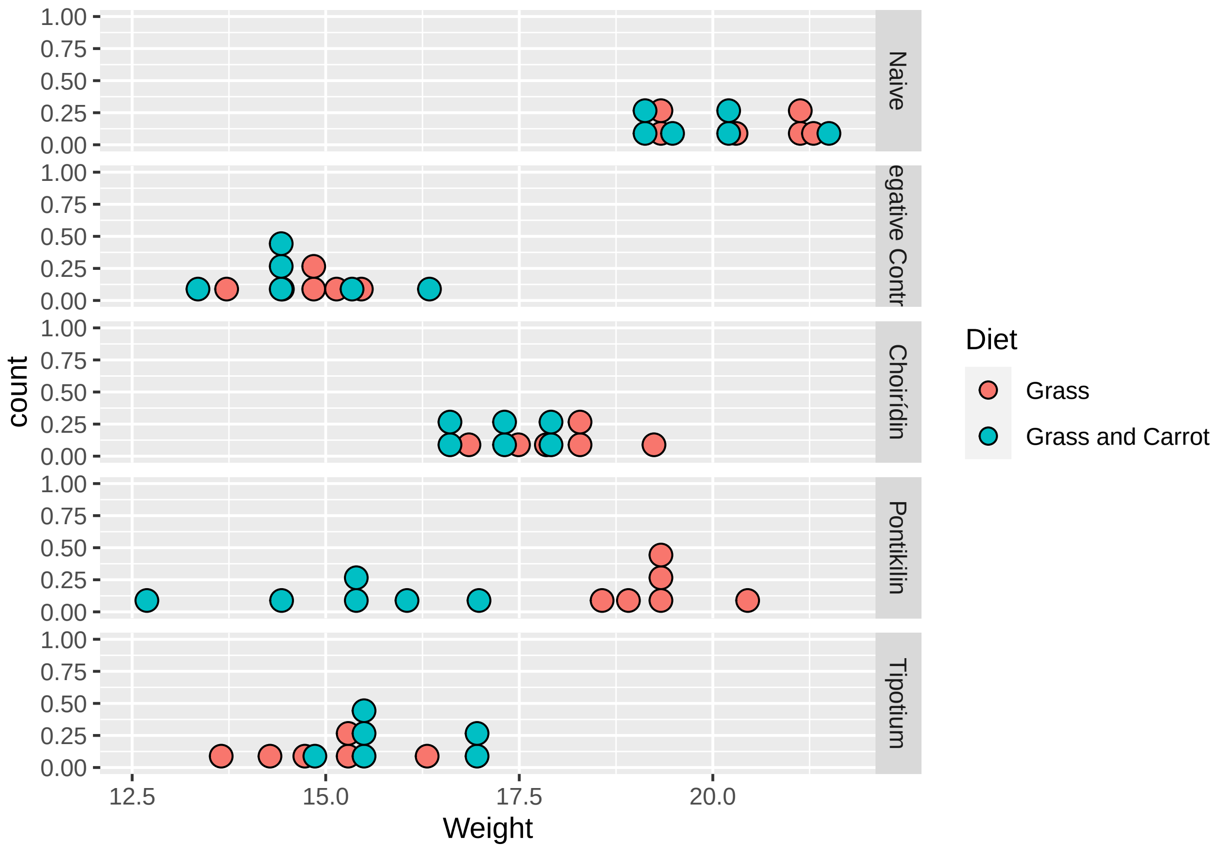

```
p3 <- ggplot(data, aes(x=Weight, y = Treatment, color = Diet)) +
  geom_boxplot(color="black", fill="white") + geom_point() +coord_flip()
p3
```




```
p4 <- ggplot(data, aes(x=Weight, fill = Diet)) +
  geom_histogram( ) + facet_grid(Treatment ~ .)
p4
```




```r
p5 <- ggplot(data, aes(x=Weight, fill = Diet)) +
  geom_dotplot() + facet_grid(Treatment ~ .)
p5
```



- Data analysis (t-Test, ANOVA with pair-wise comparisons)

ANOVA

```
data_diet1 <- data[data$Diet == "Grass",]
data_diet2 <- data[data$Diet == "Grass and Carrot",]


res.aov1 <- aov(Weight ~ Treatment, data = data_diet1)
summary(res.aov1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment    4 157.77   39.44   62.88 1.11e-12 ***
## Residuals   25  15.68    0.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res.aov2 <- aov(Weight ~ Treatment, data = data_diet2)
summary(res.aov2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment    4  106.3  26.588   25.87 1.44e-08 ***
## Residuals   25   25.7   1.028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tukey Test

```r
TukeyHSD(res.aov1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Weight ~ Treatment, data = data_diet1)
##
## $Treatment
##                                   diff         lwr        upr     p adj
## Negative Control-Naive      -5.6783333 -7.02123908 -4.3354276 0.0000000
## Choirídin-Naive             -2.4200000 -3.76290575 -1.0770943 0.0001580
## Pontikilin-Naive            -1.1083333 -2.45123908  0.2345724 0.1418758
## Tipotium-Naive              -5.4950000 -6.83790575 -4.1520943 0.0000000
## Choirídin-Negative Control   3.2583333  1.91542759  4.6012391 0.0000017
## Pontikilin-Negative Control  4.5700000  3.22709425  5.9129057 0.0000000
## Tipotium-Negative Control    0.1833333 -1.15957241  1.5262391 0.9941987
## Pontikilin-Choirídin         1.3116667 -0.03123908  2.6545724 0.0579140
## Tipotium-Choirídin          -3.0750000 -4.41790575 -1.7320943 0.0000045
## Tipotium-Pontikilin         -4.3866667 -5.72957241 -3.0437609 0.0000000
```

```r
TukeyHSD(res.aov2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Weight ~ Treatment, data = data_diet2)
##
## $Treatment
##                                   diff        lwr        upr     p adj
## Negative Control-Naive      -5.2116667 -6.9307075 -3.4926259 0.0000000
## Choirídin-Naive             -2.6650000 -4.3840408 -0.9459592 0.0010319
## Pontikilin-Naive            -4.7833333 -6.5023741 -3.0642925 0.0000002
## Tipotium-Naive              -4.0666667 -5.7857075 -2.3476259 0.0000026
## Choirídin-Negative Control   2.5466667  0.8276259  4.2657075 0.0017184
## Pontikilin-Negative Control  0.4283333 -1.2907075  2.1473741 0.9469815
## Tipotium-Negative Control    1.1450000 -0.5740408  2.8640408 0.3155836
## Pontikilin-Choirídin        -2.1183333 -3.8373741 -0.3992925 0.0104482
## Tipotium-Choirídin          -1.4016667 -3.1207075  0.3173741 0.1498497
## Tipotium-Pontikilin          0.7166667 -1.0023741  2.4357075 0.7375851
```

- Report the results: what is the best treatment available, and how confident you are that it will work. Hint: take into account the patient's preferences.

Firstly, visualizing our data we tried to compare the mean weight by diet and by treatment. We saw that there is a slight difference in the mean weight between diet groups for each treatment group. Also, we saw variety of average weights per treatment group, which was anticipated. We used boxplots to see the distribution of data in each treatment group. Clearly, we could pinpoint significant differences between the Negative Control and Naive groups, and also Negative Control with 2 of the treatments - Choirídin and Pontikilin. The difference between these 2 drug groups were the variance within. Pontikilin had large variance in the weight, while Choirídin boxplot was entirely placed above the Negative Control one. So we thought that the reason of this difference could be the diet. Visualizing the weights for each treatment group by diet distribution, we saw that Pontikilin had different effect depending on the diet.

So, as we were dealing with groups that have different diets, where diet was affecting the treatment, we supposed that the treatment for each of those groups might be different. Thus, we decided to use ANOVA test to check for impact of the drug for each diet group separately. We saw that there is a statistical difference between means of at least 2 treatment groups being examined in both diet groups with alpha ~ 0, as the p value was very small (***)

Further analysis was needed to see exactly which 2 or more groups were statistically different. So for that we used Tukey's honestly significant difference test to do pairwise comparisons. We observed the following patterns

- Tipotium did not show significant increase in body weight of treatment groups compared to control for any diet patients.
- Pontikilin showed very significant increase in body weight for Grass diet group, but no significant change in Grass and Carrot diet group.
- Choirídin showed significant difference in body weight of treatment groups compared to control for both diet groups.

We can thus conclude, that if our guinea pig has a grass diet then Pontikilin will be the best drug to use. But if he eats also carrots, or has a mixed diet, then Choirídin will be a better choice.

- Report limitations of your study

  - We can't preform experiments on guinea pigs
  - We can't preform experiment on large number of mice
  - The statistical analysis only shows us the end result (weight gain), but does not provide with any information about the working mechanism of the drug.