

HW 3

Enter your name and EID here: Harini Shanmugam hs28663

You will submit this homework assignment as a pdf file on Gradescope.

For all questions, include the R commands/functions that you used to find your answer (show R chunk). Answers without supporting code will not receive credit. Write full sentences to describe your findings.

Question 1: (2 pts)

The dataset `ChickWeight` contains information about the weights (in grams) of chicks on four different diets over time (measured at 2-day intervals) as the result of an experiment. The first few observations are listed below.

```
# Save dataset as a dataframe
ChickWeight <- as.data.frame(ChickWeight)

# Visualize the first ten rows of the dataset
head(ChickWeight,10)
```

```
##      weight Time Chick Diet
## 1       42    0     1     1
## 2       51    2     1     1
## 3       59    4     1     1
## 4       64    6     1     1
## 5       76    8     1     1
## 6       93   10     1     1
## 7      106   12     1     1
## 8      125   14     1     1
## 9      149   16     1     1
## 10     171   18     1     1
```

Use some combination of `table()` and `length()` to answer the following questions:

```
length(table(ChickWeight$Chick)) # Count number of distinct chicks are there
```

```
## [1] 50
```

```
length(table(ChickWeight$Time)) # Count number of distinct time points
```

```
## [1] 12
```

```
length(table(ChickWeight$Diet)) # Count number of distinct diet conditions
```

```
## [1] 4
```

```
# Count number of chicks per diet condition
```

```
table(ChickWeight$Time==0 & ChickWeight$Diet==1)
```

```
##
```

```
## FALSE TRUE
```

```
## 558 20
```

```
table(ChickWeight$Time==0 & ChickWeight$Diet==2)
```

```
##
```

```
## FALSE TRUE
```

```
## 568 10
```

```
table(ChickWeight$Time==0 & ChickWeight$Diet==3)
```

```
##
```

```
## FALSE TRUE
```

```
## 568 10
```

```
table(ChickWeight$Time==0 & ChickWeight$Diet==4)
```

```
##
```

```
## FALSE TRUE
```

```
## 568 10
```

- How many distinct chicks are there?
- How many distinct time points?
- How many distinct diet conditions?
- How many chicks per diet condition at the beginning of the experiment?

There are 50 distinct chicks, 12 distinct time points, and 4 distinct diet conditions.

There are ...

20 chicks with diet condition 1,

10 chicks with diet condition 2,

10 chicks with diet condition 3,

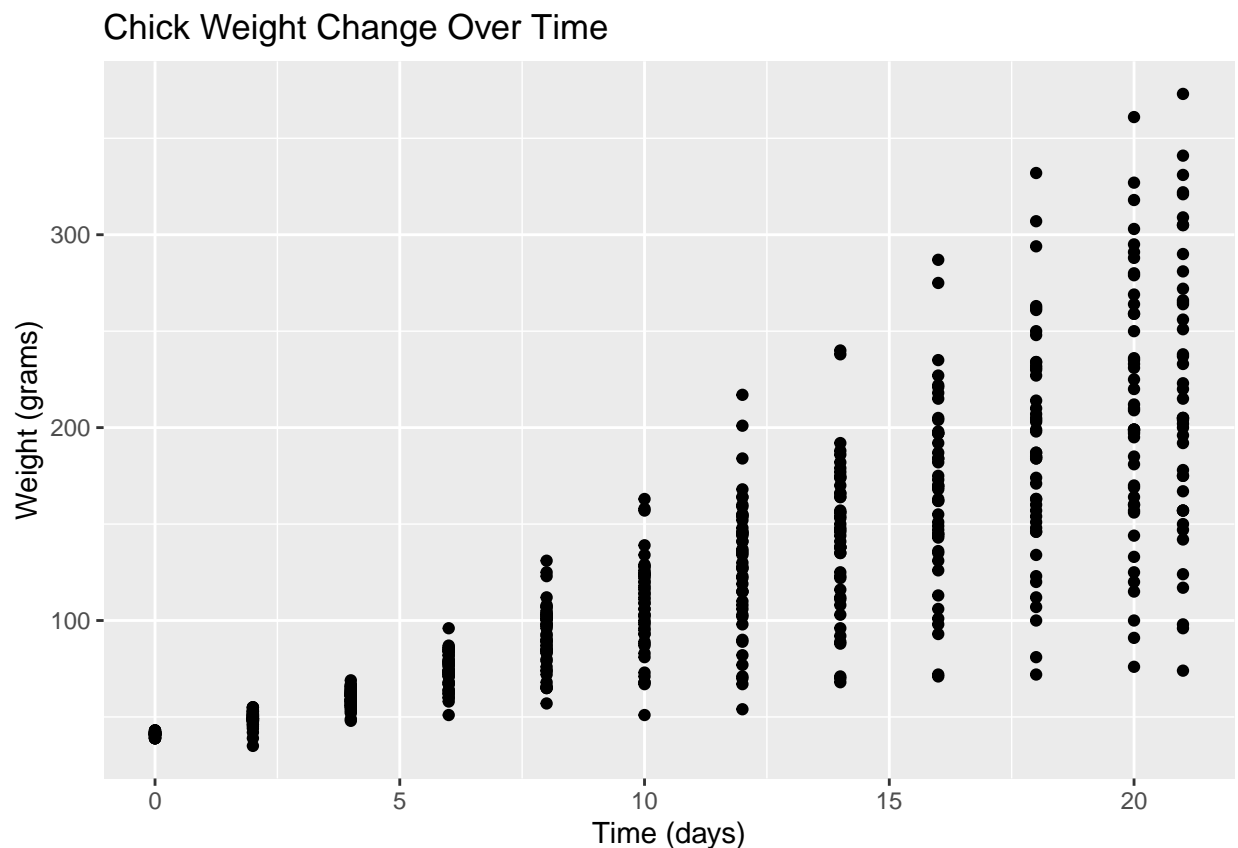
and 10 chicks with diet condition 4.

Question 2: (1 pt)

Using the `ggplot2` package, create a simple scatterplot showing chick weight (on the y-axis) depending on Time. Label the axes including the units of the variables and give the plot a title. How does chick weight change over Time?

```
# Load package
library(ggplot2)

ggplot(data = ChickWeight, aes(x=Time, y=weight)) + # Plot of Weight over Time
  geom_point() + # Add points on the plot
  labs(title = "Chick Weight Change Over Time",
       x = "Time (days)",
       y = "Weight (grams)")
```



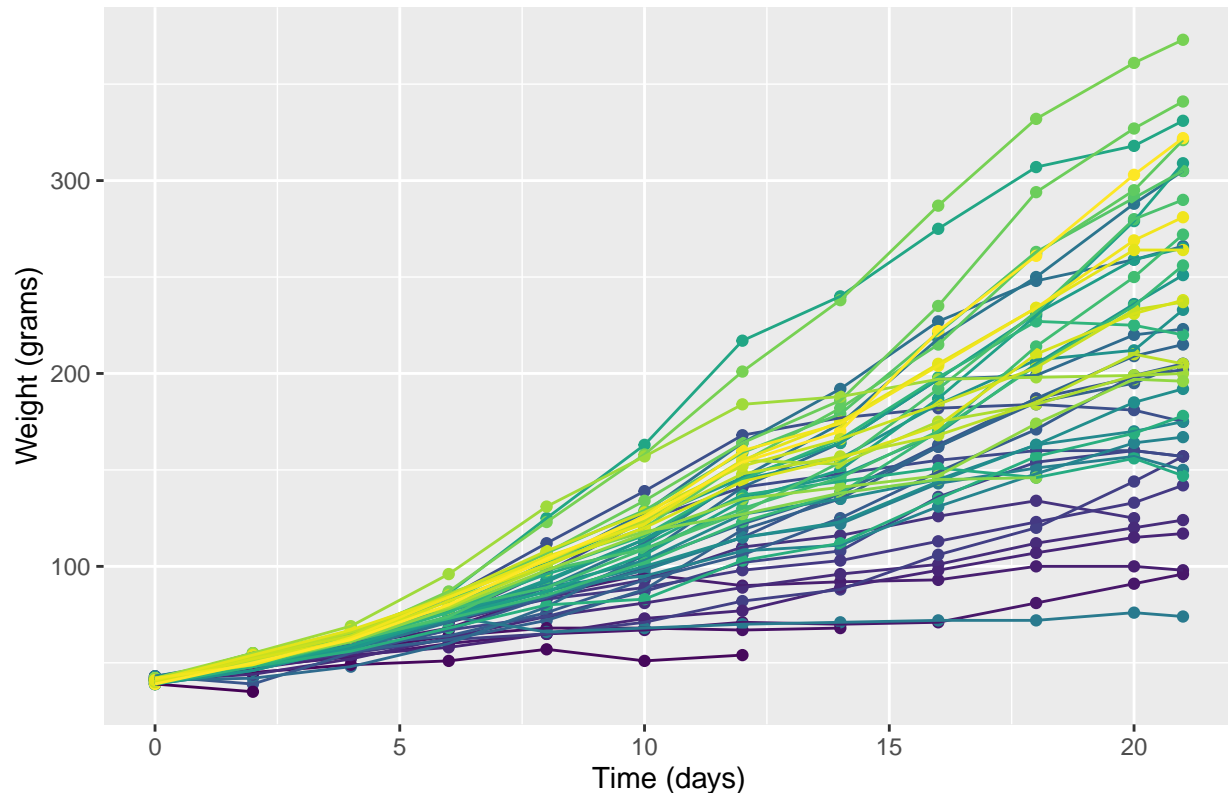
The chick weight increases over time.

Question 3: (2 pts)

Building upon the previous plot, map `Chick` to an aesthetic that assigns a color to each chick's data points. Add lines that connect each chick's points together with `geom_line()` (also colored by each chick). Finally, remove the legend. Do all chicks seem to gain weight in the same manner? Why/Why not?

```
# Plot of Weight over Time
ggplot(data = ChickWeight, aes(x=Time, y=weight, color=Chick)) +
  geom_point() + # Add points on the plot
  geom_line() + # Connect all points on the plot for each chick
  labs(title = "Chick Weight Change Over Time",
       x = "Time (days)",
       y = "Weight (grams)") +
  theme(legend.position = "none") # Get rid of the legend
```

Chick Weight Change Over Time



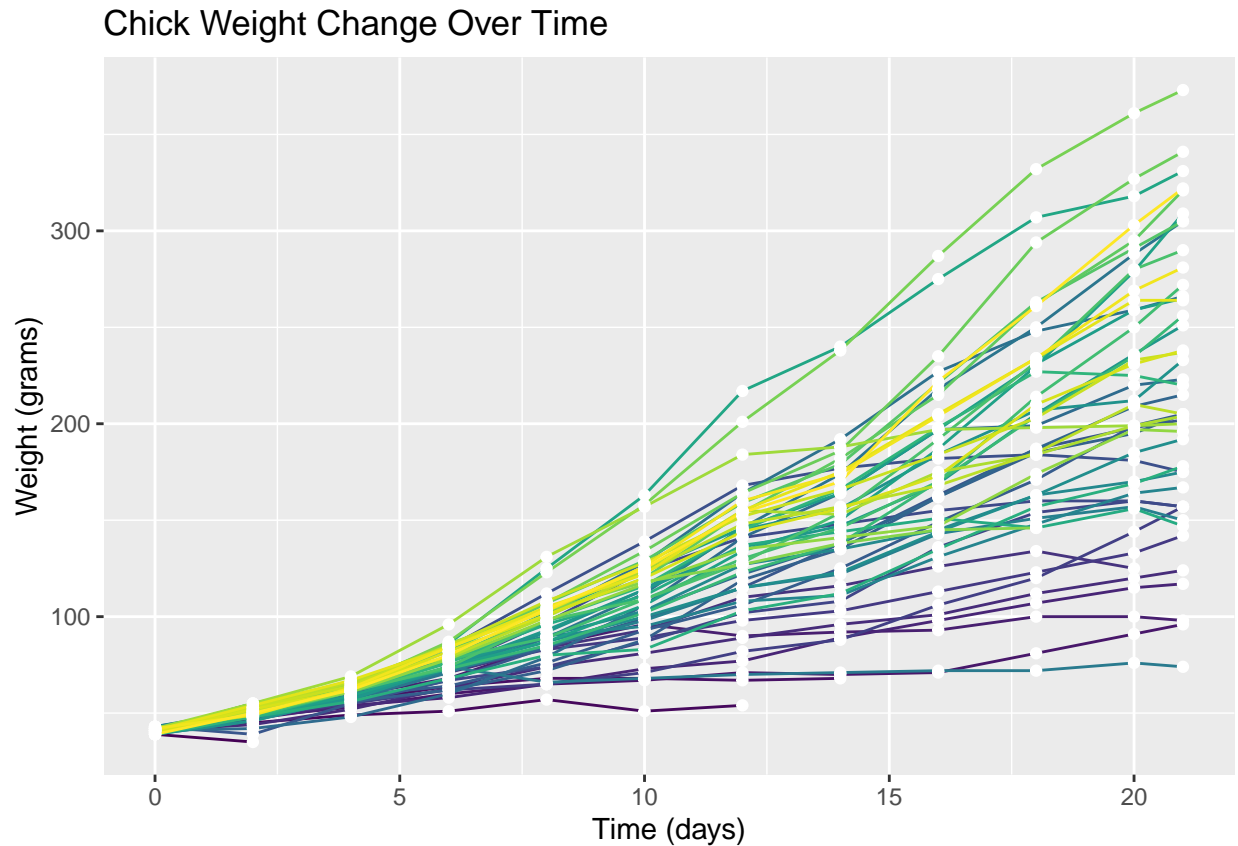
No, not all chicks gain weight in the same manner. Some only keep gaining weight and at faster and faster rate. For others, they slow down the rate at which they gain weight and even start to lose some weight.

Question 4: (1 pt)

Continue modifying the same graph by removing the color from the points only: leave the lines colored by chick, but make all of the points white. Make sure to put the points *on top of* the lines. On which day was the last value of the chicks' weight recorded?

```
# Plot of Weight over Time colored by each chick
ggplot(data = ChickWeight, aes(x=Time, y=weight, color=Chick)) +
  geom_line() + # Connect points on the plot by each chick
```

```
geom_point(color="white") + # Add points on the plot and make them white
labs(title = "Chick Weight Change Over Time",
     x = "Time (days)",
     y = "Weight (grams)") +
theme(legend.position = "none") # Get rid of the legend
```



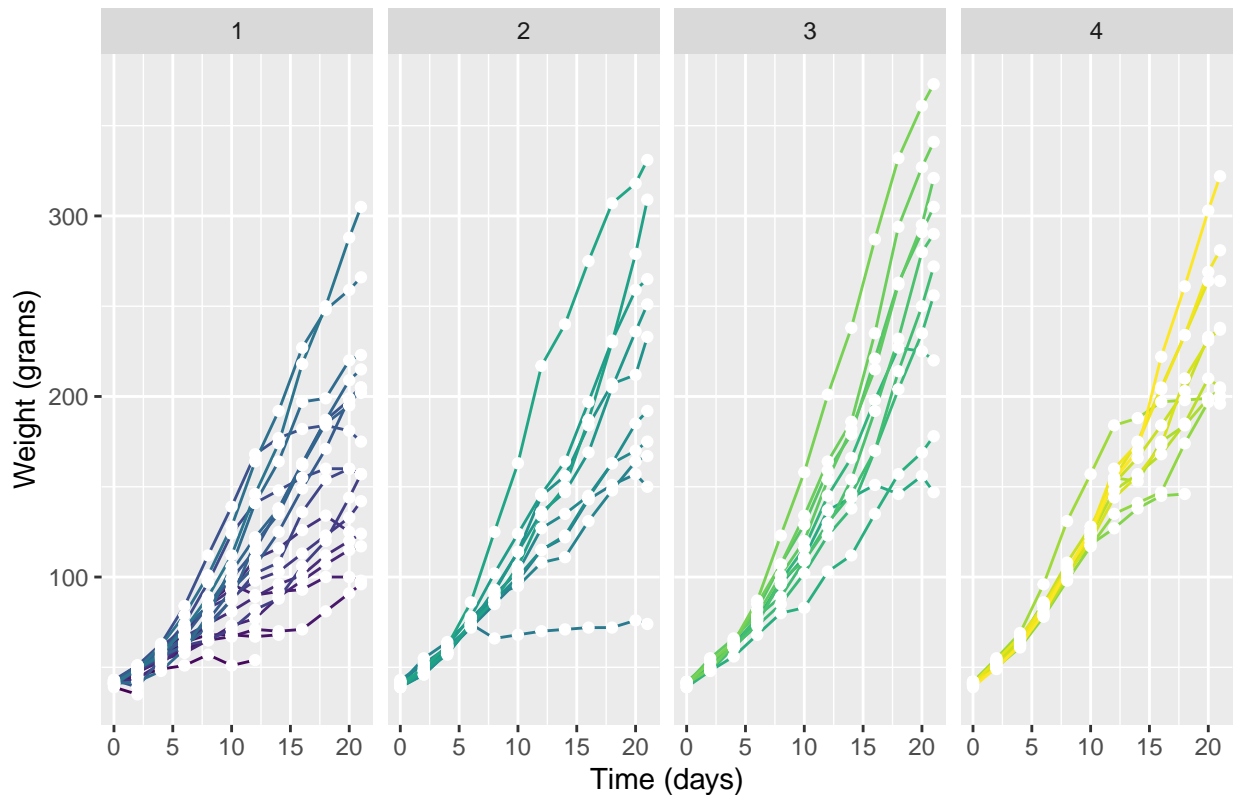
The last value of the chicks' weight was recorded on Day 21.

Question 5: (2 pts)

Now, facet this plot by diet. Can you tell from this new plot which diet results in greater weight? Describe how the relationship between weight and Time changes, or not, across the different diets.

```
# Plot of Weight over Time colored by each chick
ggplot(data = ChickWeight, aes(x=Time, y=weight, color=Chick)) +
  geom_line() + # Connect points on the plot by each chick
  geom_point(color="white") + # Add points on the plot and make them white
  labs(title = "Chick Weight Change Over Time by Diet",
       x = "Time (days)",
       y = "Weight (grams)") +
  theme(legend.position = "none") + # Get rid of the legend
  facet_grid(~Diet) # Split into separate graphs by diet
```

Chick Weight Change Over Time by Diet



Yes, we can tell which diet results in greater weight. Diet 3 recorded the greatest weight. Diet 2 recorded the second highest weights but had the largest spread. Diet 4 had the tightest spread of weights.

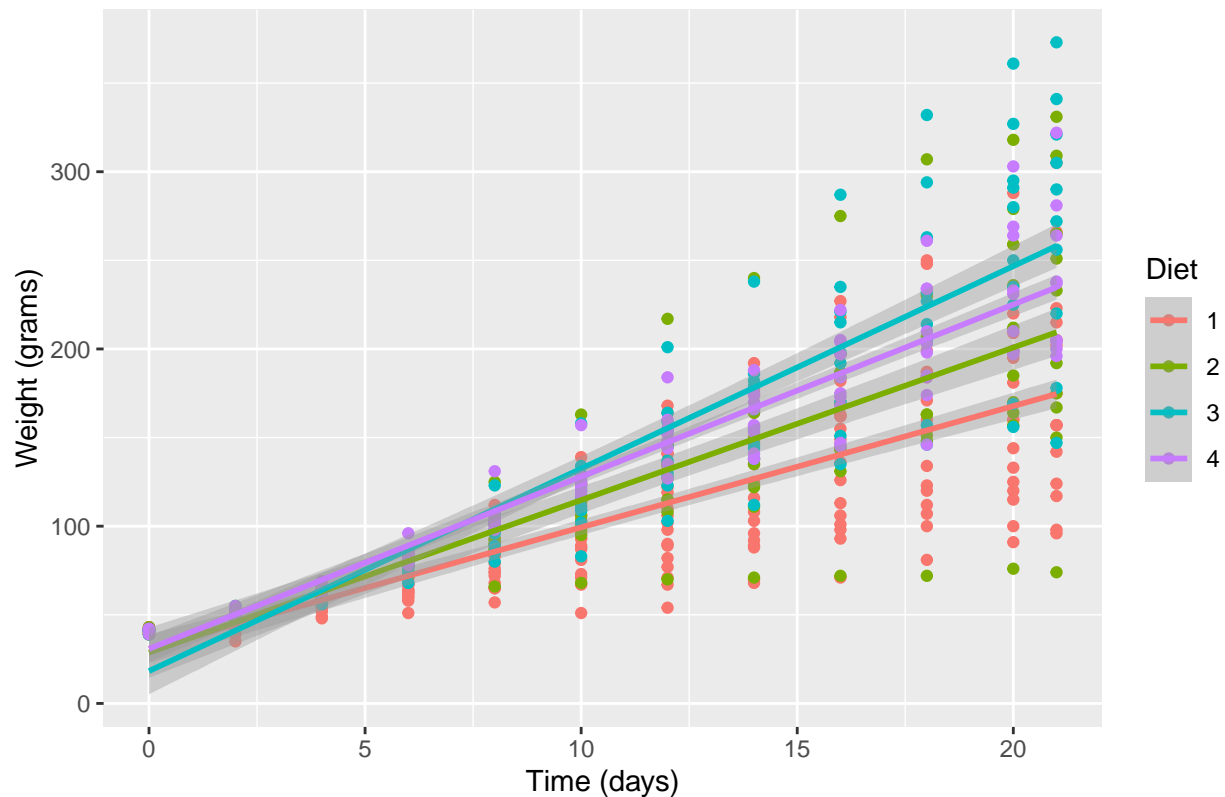
The relationship between weight and change do not really change. Across all 4 diets/graphs, as time increases, weight increases.

Question 6: (2 pts)

Go back to your plot from question 2 and fit a *linear regression line* (using `method = "lm"` in `geom_smooth()`) to the chicks in each diet with `geom_smooth()`. There should be 4 separate regression lines, one for each diet, each a separate color. Can you see more clearly which diet results in greater weight? Explain.

```
# Plot of Weight over Time for each diet
ggplot(data = ChickWeight, aes(x=Time, y=weight, color=Diet)) +
  geom_point() + # Add points on the plot
  labs(title = "Chick Weight Change Over Time",
       x = "Time (days)",
       y = "Weight (grams)") +
  geom_smooth(method = "lm") # Linear regression line
```

Chick Weight Change Over Time



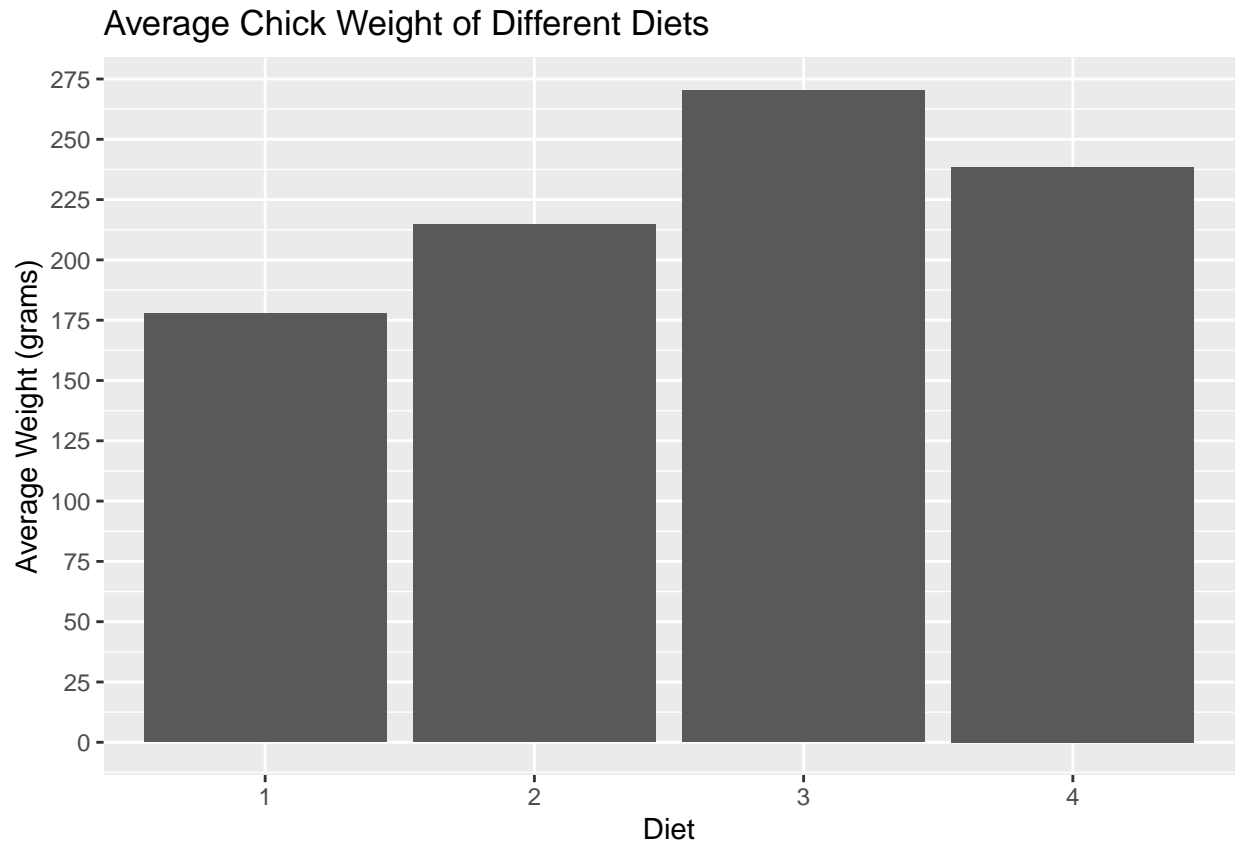
Yes, it is clearer as to which diet results in greater weight. Diet 3 results in greater weight (has largest slope/ fastest rate of weight gain).

Question 7: (2 pts)

A scatterplot might not be the best way to visualize this data: it calls attention to the relationship between weight and time, but it can be hard to see the differences between diets. A more traditional approach for exploring the effect of diet would be to construct a bar graph representing group means at the end of the experiment.

Index the data in the `ggplot` function to focus on the last Time point. *Hint: Refer to Question 4 or find the max Time*). Then create a plot using `geom_bar` where each bar's height corresponds to the average chick weight for each of the four diet conditions. Rename the y-axis to include units (e.g., with `scale_y_continuous(name=...)`) and make the major tick marks go from 0 to 300 by 25 (e.g., with `scale_y_continuous(breaks=...)`). Which diet has the highest mean weight?

```
# Index for only max chick weight
ggplot(data=ChickWeight[ChickWeight$Time==21,], aes(x=Diet, y=weight)) +
  geom_bar(stat="summary", fun="mean") + # Bar graph of avg weight for each diet
  scale_y_continuous(name="Average Weight (grams)", # Y-Axis title rename
    breaks=seq(0,300,25)) + # Ticks 0-300 by increments of 25
  labs(title="Average Chick Weight of Different Diets")
```

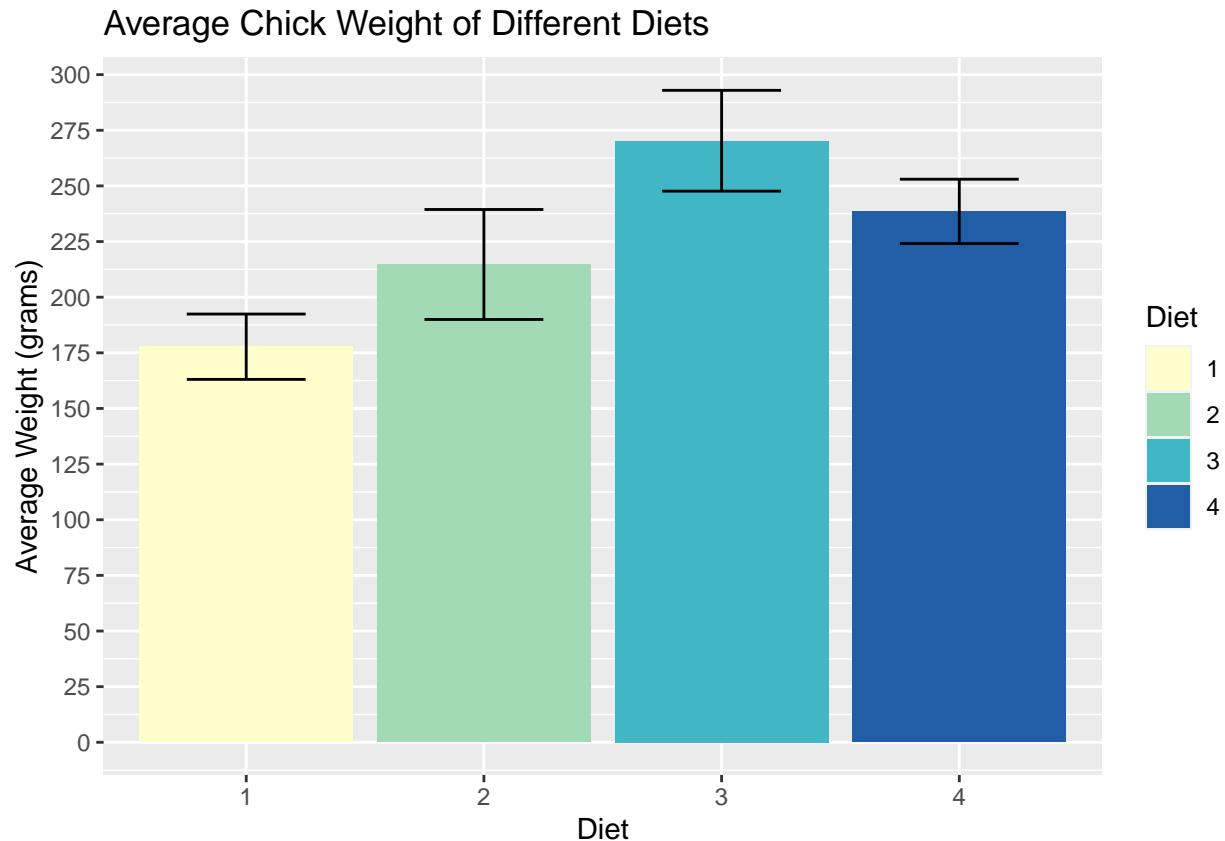


Diet 3 has the highest mean weight.

Question 8: (2 pts)

Building on the previous graph, add error bars showing ± 1 SE using `geom_errorbar(stat = "summary")`. Make the error-bars skinnier by adding a `width = 0.5` argument. Color the bars (not the error bars, but the bar graph bars) by diet and change from the default color scheme using a `scale_fill_` or a `scale_color_`. Compare the different diets in terms of variation in weight.

```
# Index for only max chick weight
ggplot(data=ChickWeight[ChickWeight$Time==21,], aes(x=Diet, y=weight, fill=Diet)) +
  geom_bar(stat="summary", fun="mean") + # Bar graph of avg weight for each diet
  geom_errorbar(stat = "summary", fun.data = "mean_se", width=0.5) +
  scale_y_continuous(name="Average Weight (grams)", # Y-Axis title rename
                     breaks=seq(0,300,25)) + # Ticks 0-300 by increments of 25
  labs(title="Average Chick Weight of Different Diets") +
  scale_fill_brewer(palette="YlGnBu") # Color bars of the graph
```

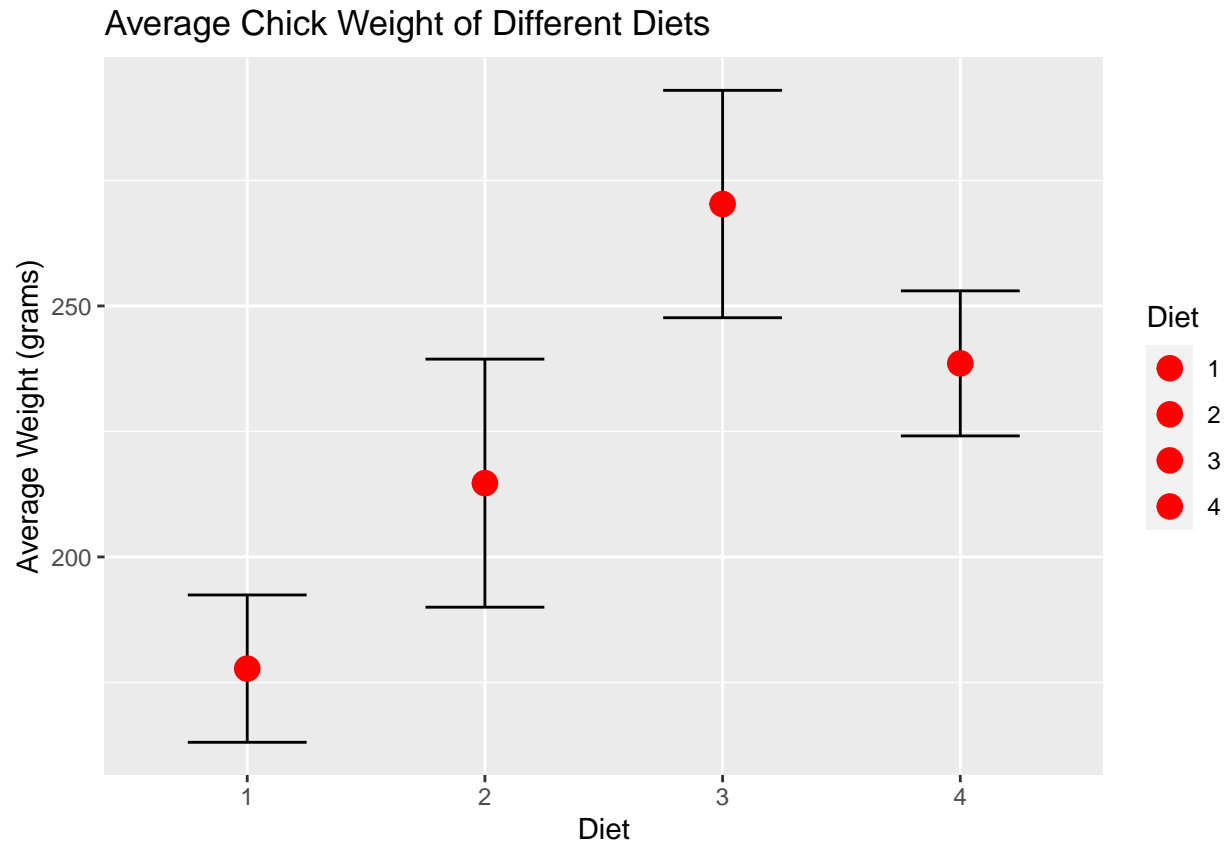



Diets 2 has the largest variation, followed by Diet 3 with a similarly large variation, then Diet1, and then Diet 4 with the smallest variation.

Question 9: (2 pts)

Take your code from question 8 and replace `geom_bar()` with `geom_point()`. Remove the `breaks =` argument from `scale_y_continuous`. Make the points larger and color them all red. Put them *on top of* the error bars. Does the mean chick weight seem to differ based on the diet? *Note: Informally state if they seem to differ and if so, how.*

```
# Index for only max chick weight
ggplot(data=ChickWeight[ChickWeight$Time==21,], aes(x=Diet, y=weight, fill=Diet)) +
  geom_errorbar(stat = "summary", fun.data = "mean_se", width=0.5) +
  # Avg weight for each diet as a red dot
  geom_point(stat="summary", fun="mean", size=4, color="Red") +
  scale_y_continuous(name="Average Weight (grams)") + # Y-Axis title rename
  labs(title="Average Chick Weight of Different Diets")
```



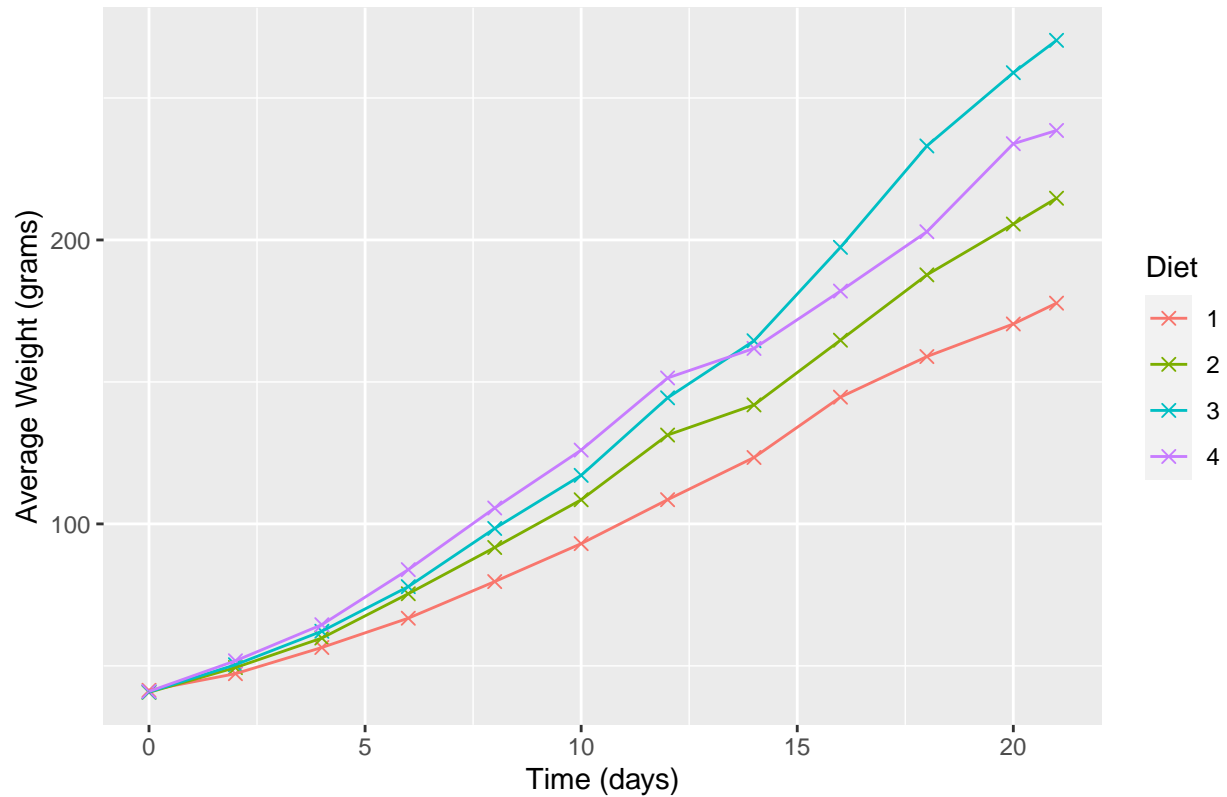
Yes, the mean chick weight seem to differ based on the diet. The mean weight of Diet 3 is the highest followed by Diet 4, Diet 2, and then Diet 1 with the smallest average weight.

Question 10: (2 pts)

One last graph! And a little less guided... It would be even more meaningful to compare the mean weight of each Diet over Time! Use `geom_point` to represent the mean weight over time and `geom_line` to connect the mean weights per diet. Change the shape of the points to be x symbols. *Giving you a hint anyway: color by diet and use some `stat` options in the geoms!* Which diet has a constantly lower mean weight over time?

```
ggplot(data=ChickWeight, aes(x=Time, y=weight, color=Diet)) +
  geom_point(stat="summary", fun="mean", size=2, pch=4) + # Points as 'x'
  geom_line(stat="summary", fun="mean") + # Mean lines to connect points
  labs(title= "Average Chick Weight Over Time by Different Diets",
        x="Time (days)",
        y="Average Weight (grams)")
```

Average Chick Weight Over Time by Different Diets



Diet 1 has a constantly lower mean weight over time.

Formatting: (2 pts)

Comment your code, write full sentences, and knit your file!

```
## sysn
## "Darw
## rele
## "21.6
## vers
## "Darwin Kernel Version 21.6.0: Wed Aug 10 14:28:35 PDT 2022; root:xnu-8020.141.5~2/RELEASE_ARM64_T81
## noden
## "Harinis-Air.attlocal.n
## mach
## "arm
## lo
## "ro
## us
## "harinishanmug
```


##

effective_u
"harinishanmug