DataVisualization

Kalen Clifton

8/4/2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

To run each chunk, click on the green arrow button in each code block.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

Set up for Data Visualization

Uncomment and run chunk to install the package ggplot2. Recomment out this section before knitting the document

```
#install.packages("ggplot2")
```

Open ggplot2 library

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

Plotting Example

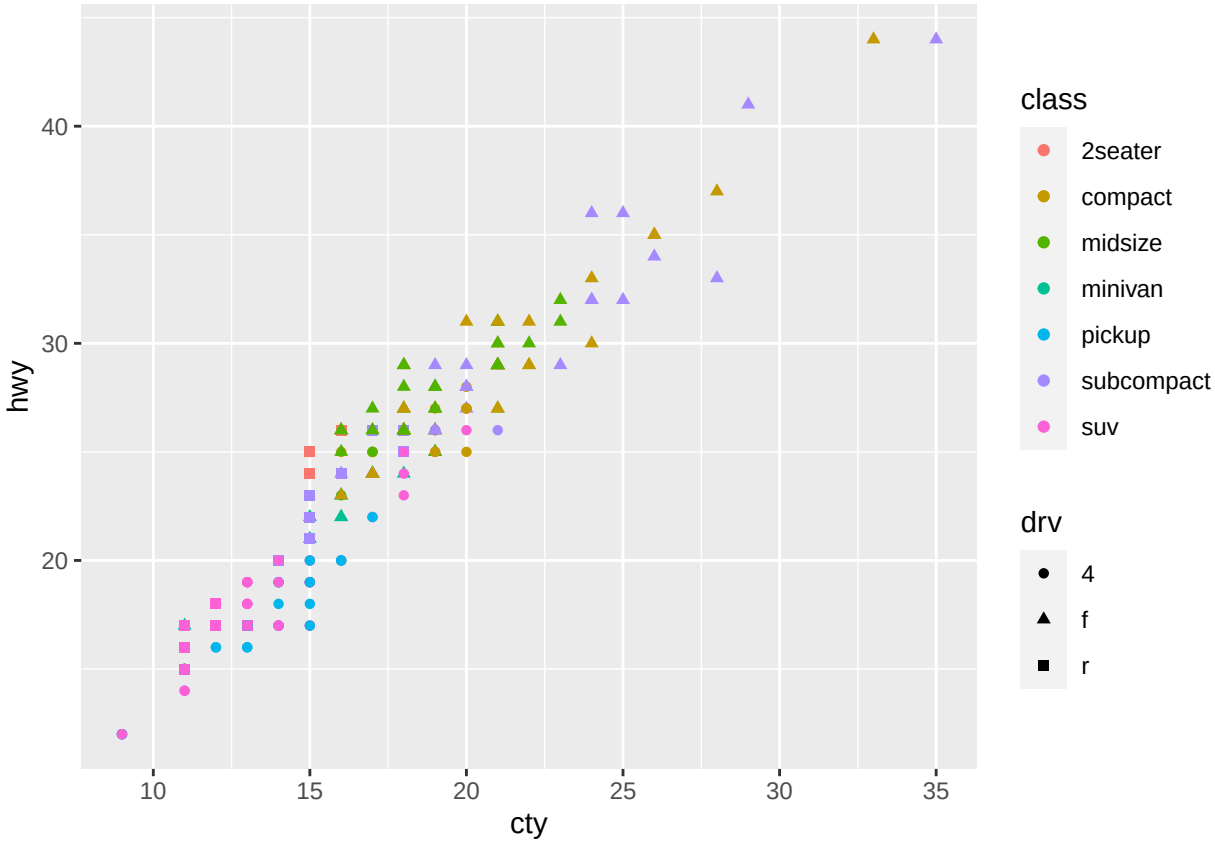
The dataset `mpg` has data on several car models collected the US EPA. The plot below displays some of that data. The variable on the x-axis is `cty`, a car's fuel efficiency in the city, in miles per gallon (mpg). The variable on the y-axis is `hwy`, a car's fuel efficiency on the highway in mpg.

Two other variables are included in the plot by aesthetic mapping. An aesthetic is a visual property of the objects in your plot. Aesthetics include things like the size, the shape, or the color of your points.

The `class` variable is added to the plot by mapping the levels of `color` to the values of `class`. (The `class` variable of the `mpg` dataset classifies cars into groups such as compact, midsize, and SUV.)

The `drv` variable (identifies type of drive including four wheel drive, front-wheel drive, and rear wheel drive) is added to the plot by mapping the levels of `shape` to the values of `drv`.

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = cty, y = hwy, color = class, shape = drv))
```



Code Template. If desired use for prompts below, replace the bracketed sections and delete the #'s

```
# ggplot(data = <DATA>) +
#  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Visualizing Data from Experiments on Cyanobacteria Growth

1. Load in `growth` data from characterization team (artificial data for example only)

```
growth <- read.csv('https://raw.githubusercontent.com/kpclifton/Baltimore_BioCrew/master/app/examples/Da
```

2. View how the data is structured by printing the first 10 rows.

```
head(growth, 10)
```

```
##    Days Iron   OD Trial
## 1     1   0x 0.05   1st
## 2     2   0x 0.05   1st
## 3     3   0x 0.05   1st
## 4     4   0x 0.05   1st
## 5     5   0x 0.05   1st
## 6     6   0x 0.05   1st
## 7     7   0x 0.05   1st
## 8     8   0x 0.05   1st
## 9     9   0x 0.05   1st
## 10   10   0x 0.05   1st
```

3. View a summary of the data to understand basic statistics about the values in the columns (such as counts for categorical variables and means for continous variables).

```
summary(growth)
```

```
##       Days          Iron          OD          Trial
##  Min.   : 1.0   0x    :40   Min.   :0.0475   1st:50
##  1st Qu.: 3.0   1/100x:40   1st Qu.:0.0550   2nd:50
##  Median : 5.5   1/10x :40   Median :0.6150   3rd:50
##  Mean   : 5.5   10x   :40   Mean   :1.0926   4th:50
##  3rd Qu.: 8.0   1x    :40   3rd Qu.:2.0000
##  Max.   :10.0               Max.   :3.3000
```

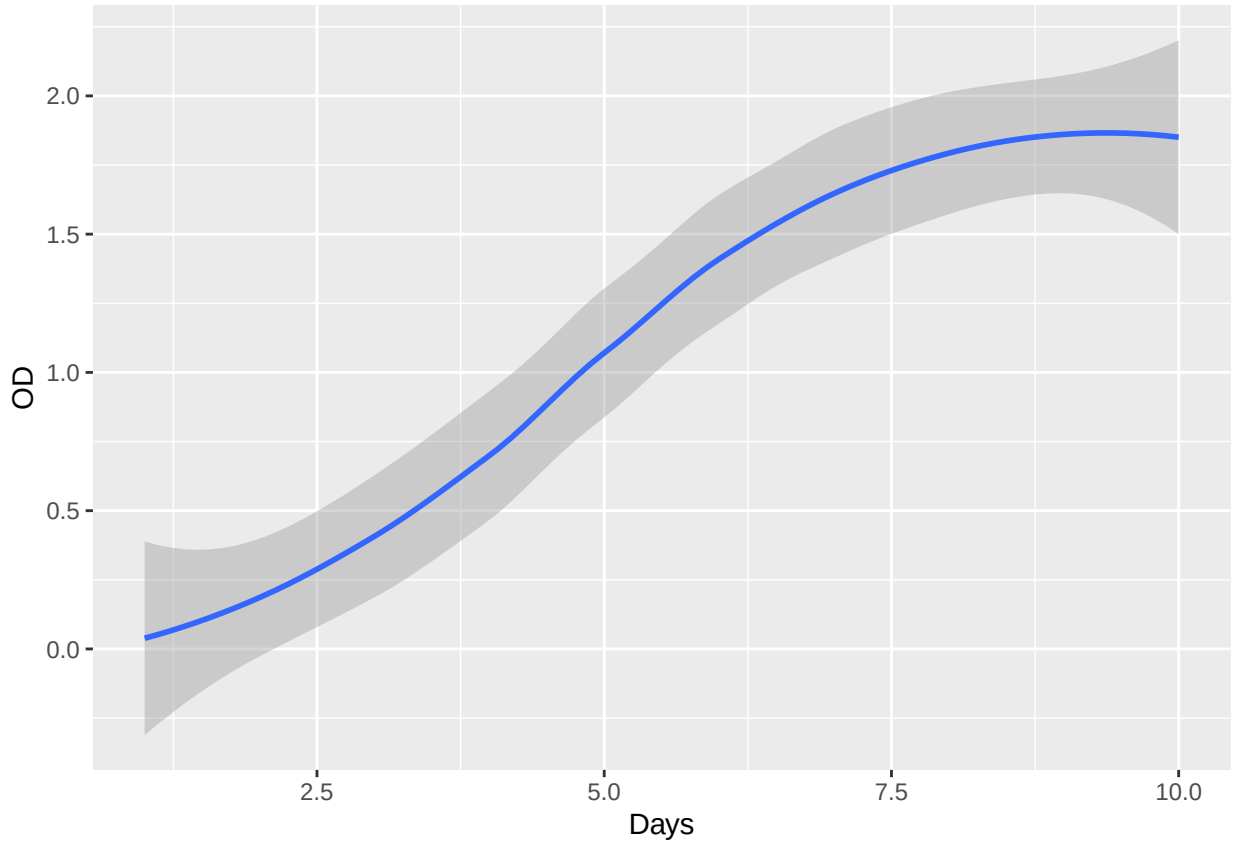
Optional: The `View` function opens the dataframe in another tab where you can scroll through the whole dataset

```
#View(growth)
```

4. Plot `Days` vs `OD` as smooth line using `geom_smooth`

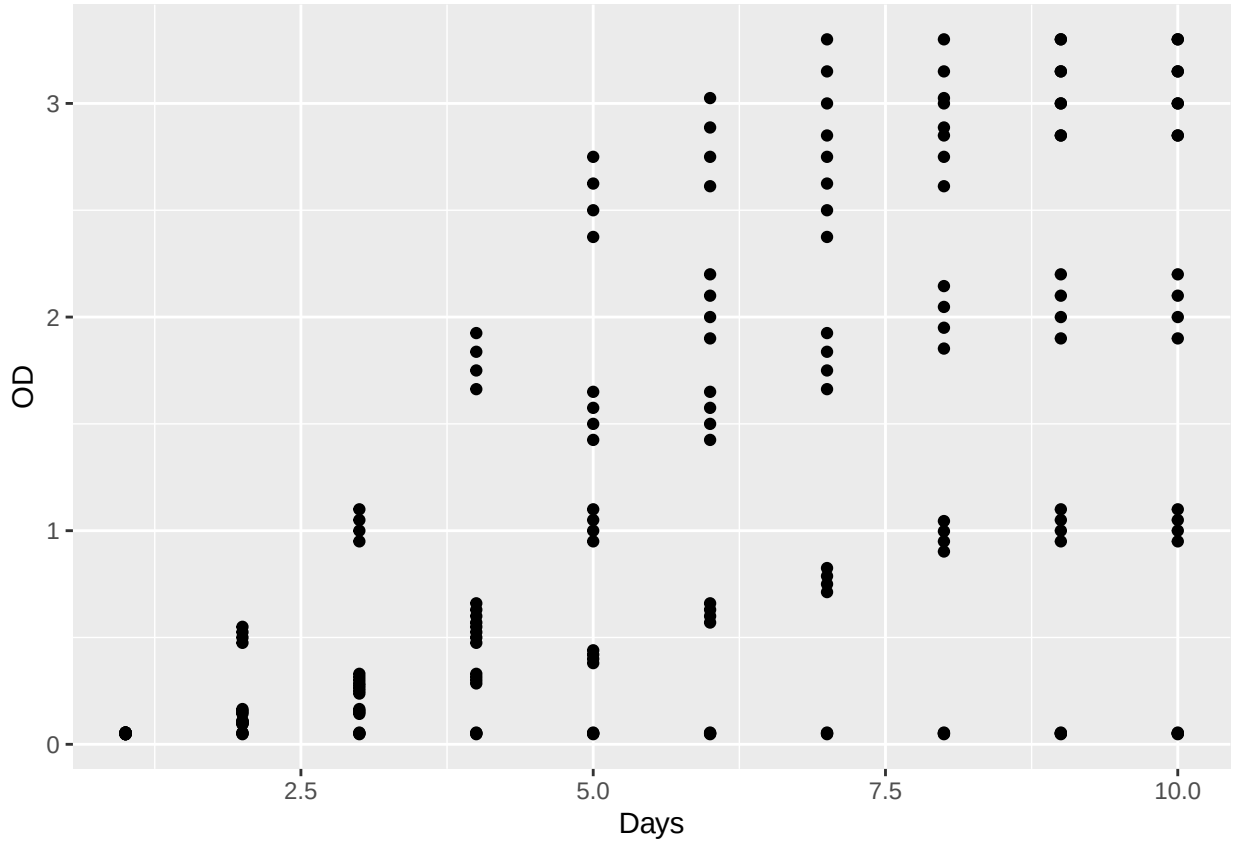
```
ggplot(data = growth) +
  geom_smooth(mapping = aes(x = Days, y = OD))
```

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



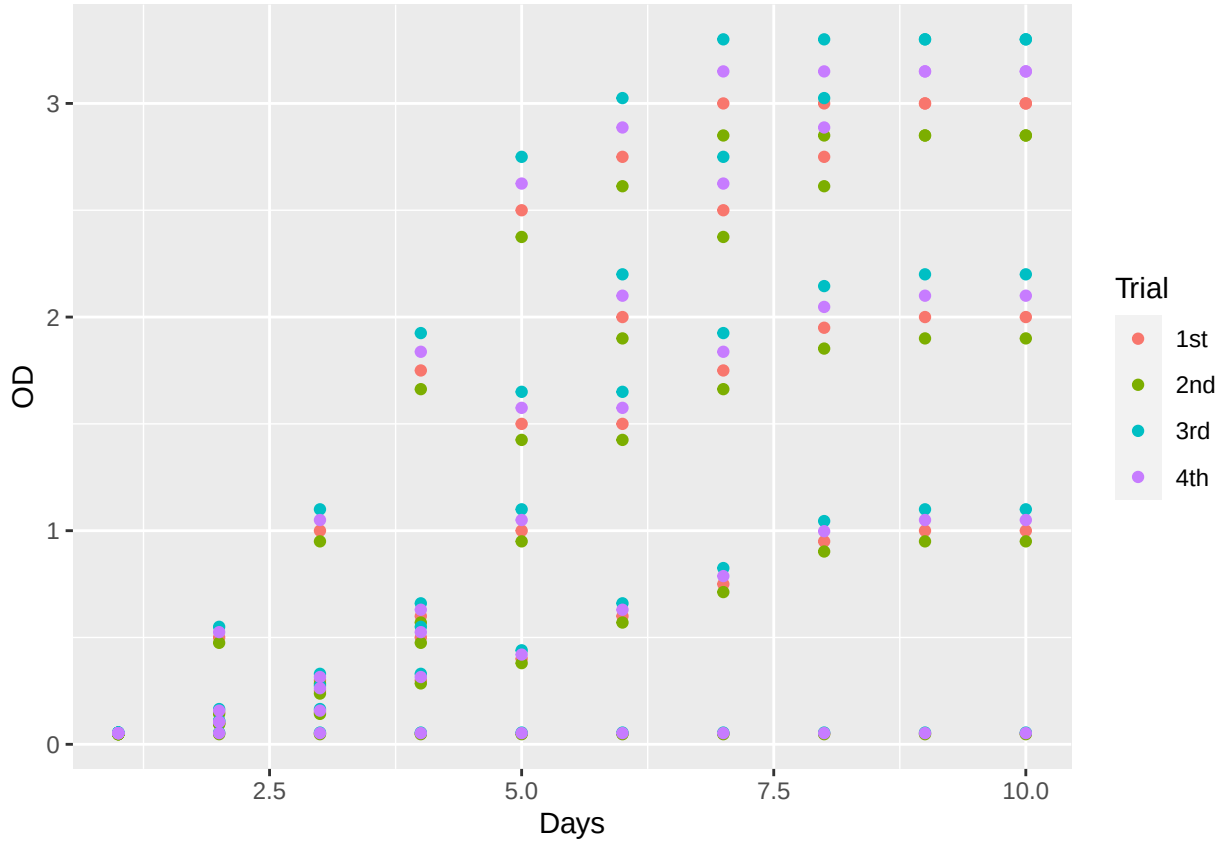
5. Plot `Days` vs `OD` as scatter plot using `geom_point`

```
ggplot(data = growth) +
  geom_point(mapping = aes(x = Days, y = OD))
```



6. Plot `Days` vs `OD` as scatter plot and add a label by mapping the variable `Trial` to `color`

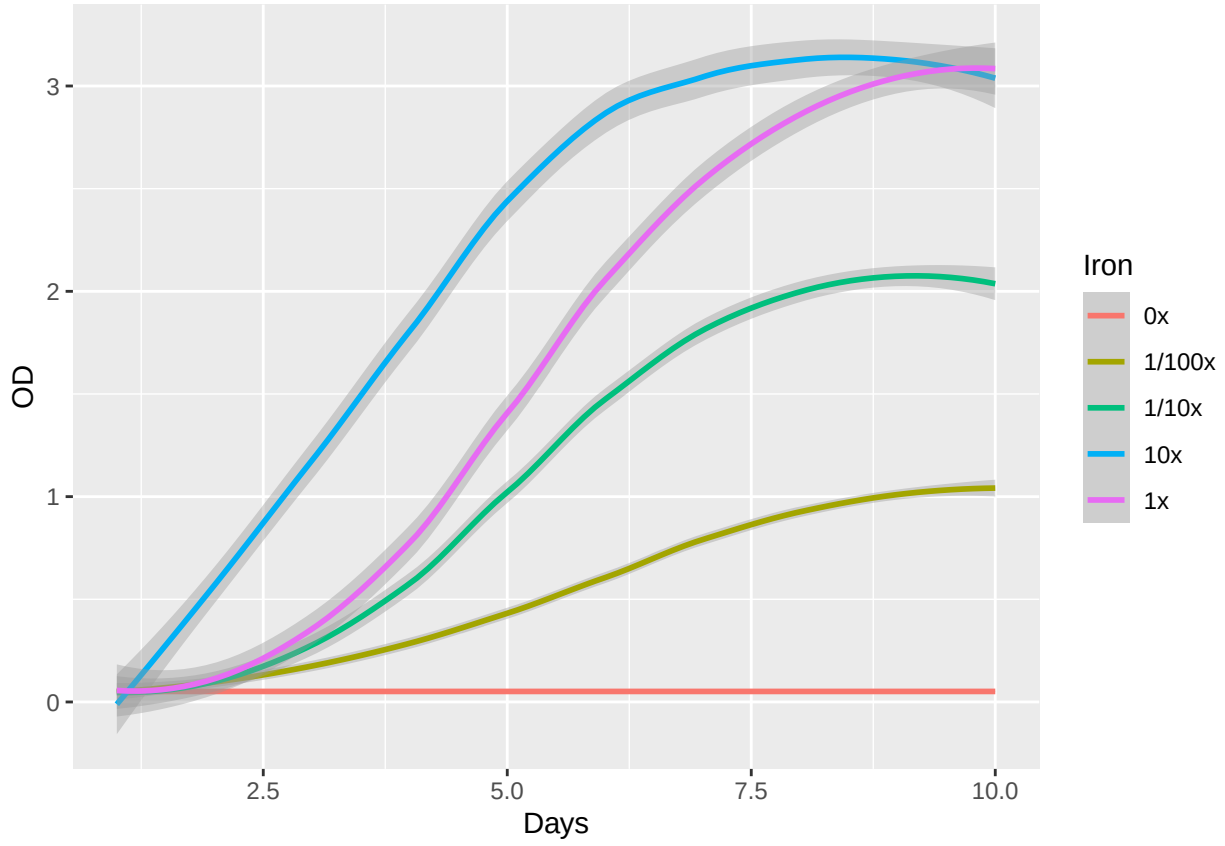
```
ggplot(data = growth) +
  geom_point(mapping = aes(x = Days, y = OD, color = Trial))
```



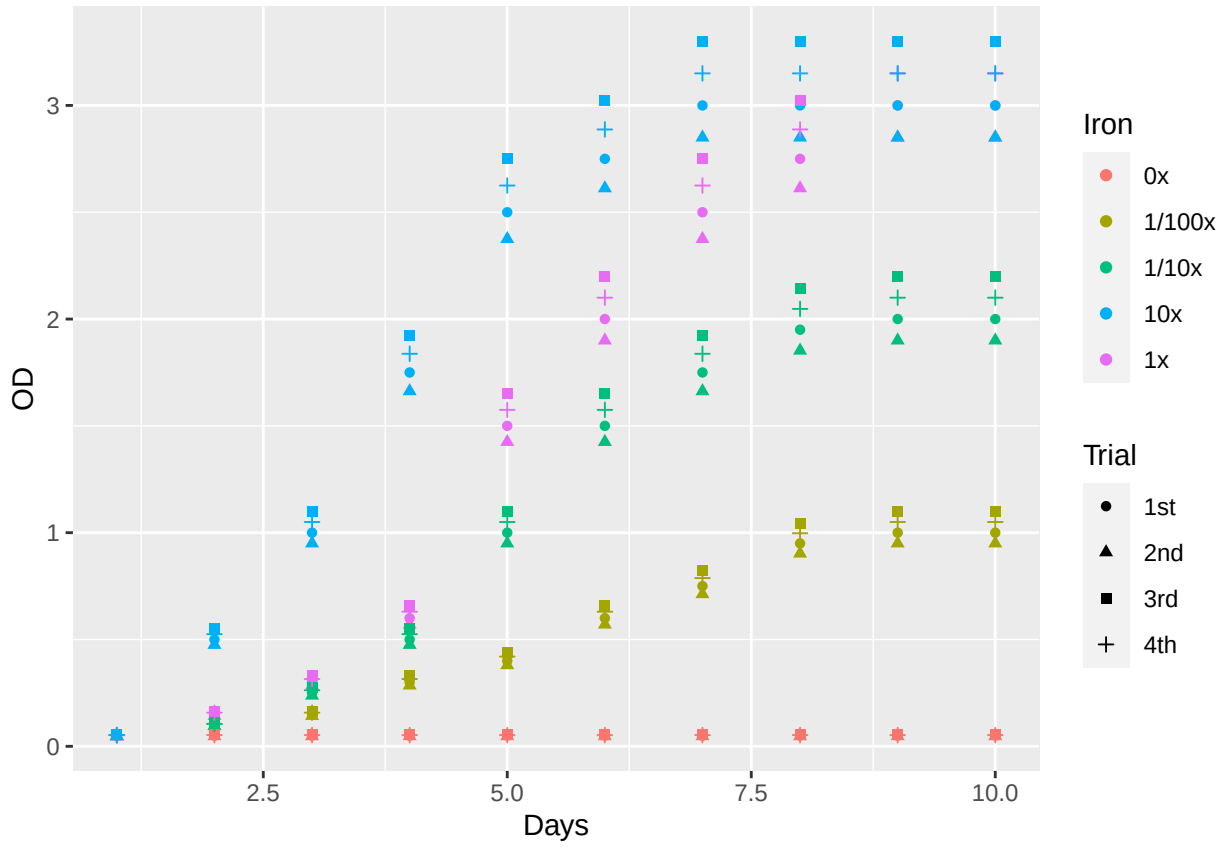
7. Try to create new plot(s) of your own design. Consider what you think would be useful for the project or simply test out different different functions and settings. You can use the cheat sheet at https://rstudio.com/wp-content/uploads/2016/11/ggplot2-cheatsheet-2.1.pdf

```
ggplot(data = growth) +
  geom_smooth(mapping = aes(x = Days, y = OD, color = Iron))
```

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




```
ggplot(data = growth) +
  geom_point(mapping = aes(x = Days, y = OD, color = Iron, shape = Trial))
```


OD
3
2
1
0
2.5
5.0
7.5
10.0
Days
Iron
0x
1/100x
1/10x
10x
1x
Trial
1st
2nd
3rd
4th