Lab 01 - Hello R

Nathan Kim

1-16-20. Due Mon, Jan 20 at 11:59p

Load Packages

```
library(tidyverse)
library(datasauRus)
```

Exercise 1

(Type your answer to Exercise 1 here. This exercise does not require any R code). **Remember that in the lab we are telling you when to stage, commit, and push**!

There are 1846 rows and 3 columns/variables. The variables are dataset (indicates which dataset the data are from), x (x-values), and y (y-values).

When turning a document in on Gradescope, remember to knit to .pdf and turn in that .pdf document.

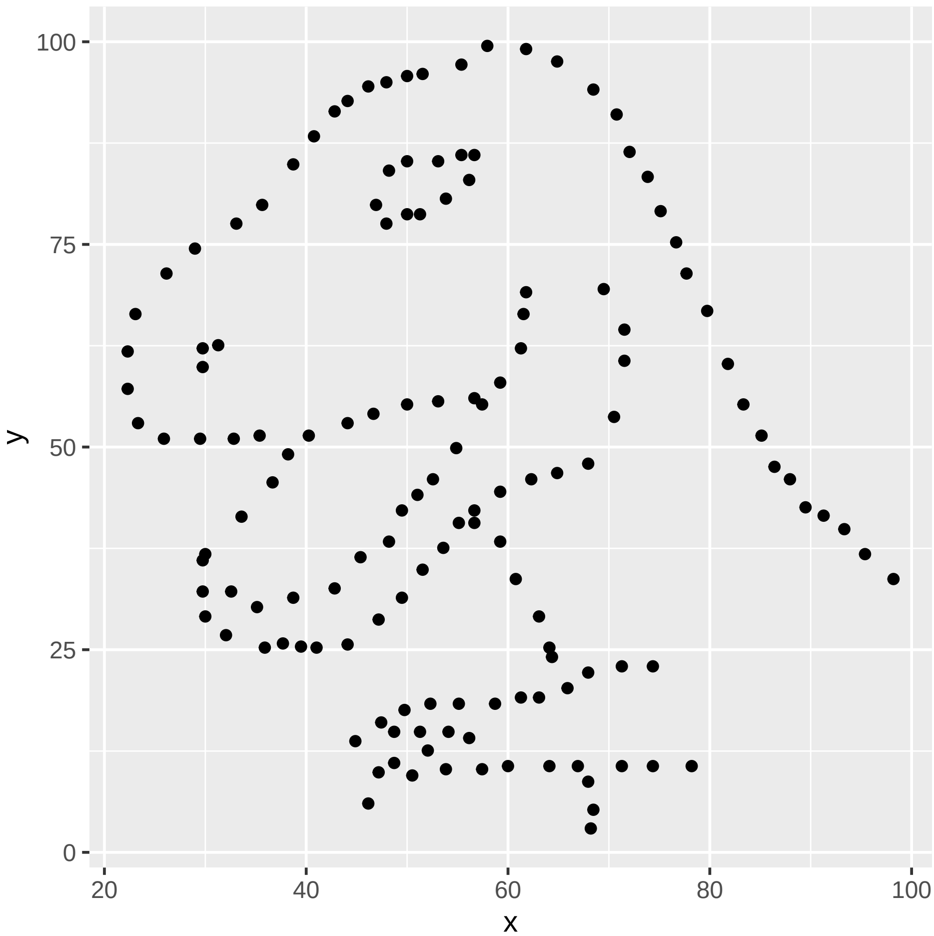
Exercise 2

(The answers for this Exercise are given for you below, but you should clean up some of the narrative so that it only includes what you want to turn in)

First, let's plot the data in the dino dataset (we covered ggplot2 in class on Wednesday):

```
dino_data <- datasaurus_dozen %>%
  filter(dataset == "dino")

ggplot(data = dino_data, mapping = aes(x = x, y = y)) +
  geom_point()
```



And next calculate the correlation between `x` and `y` in this dataset (an explanation for this code is given in the lab document. We will cover data wrangling in more detail in class next Wednesday):

```
dino_data %>%
  summarize(r = cor(x, y))
```

```
## # A tibble: 1 x 1
##         r
##     <dbl>
## 1 -0.0645
```

Exercise 3

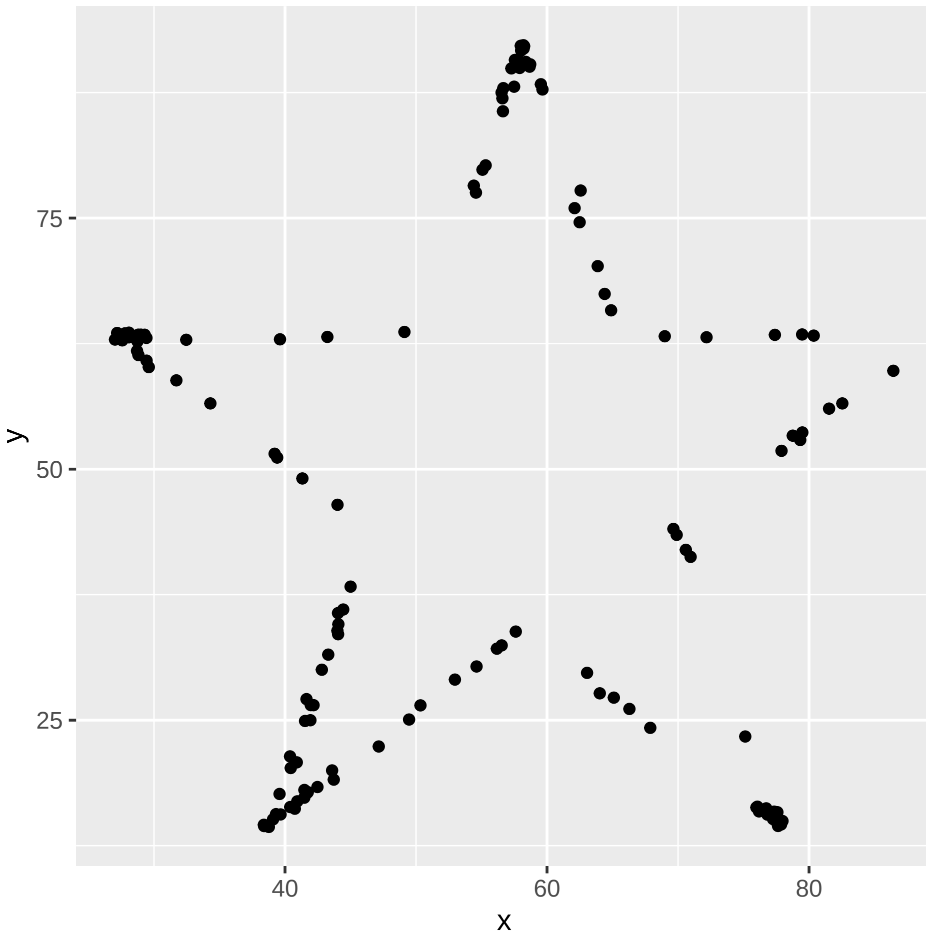
(Add code and narrative as needed. Note that the `R` chunks are labeled with `plot-star` and `cor-star` to provide spaces to place the code for plotting and calculating the correlation coefficient. To finish, clean up the narrative by removing these instructions)

The correlation coefficient for star_data is -0.0629611 as opposed to dino_data's -0.06447185. This ultimately does not mean much as both closely follow a shape

```
star_data <- datasaurus_dozen %>%
  filter(dataset == "star")

ggplot(data = star_data, mapping = aes(x = x, y = y)) +
```

```
  geom_point()
```



We've filtered the data to only include the points that are under the dataset 'star', and then assigning those data points into the name star_data. Then we plotted the points under the dataset 'star' using ggplot. This created a scatterplot that creates the shape of a star, rather than a dinosaur.

```
star_data %>%
  summarize(r = cor(x, y))
```

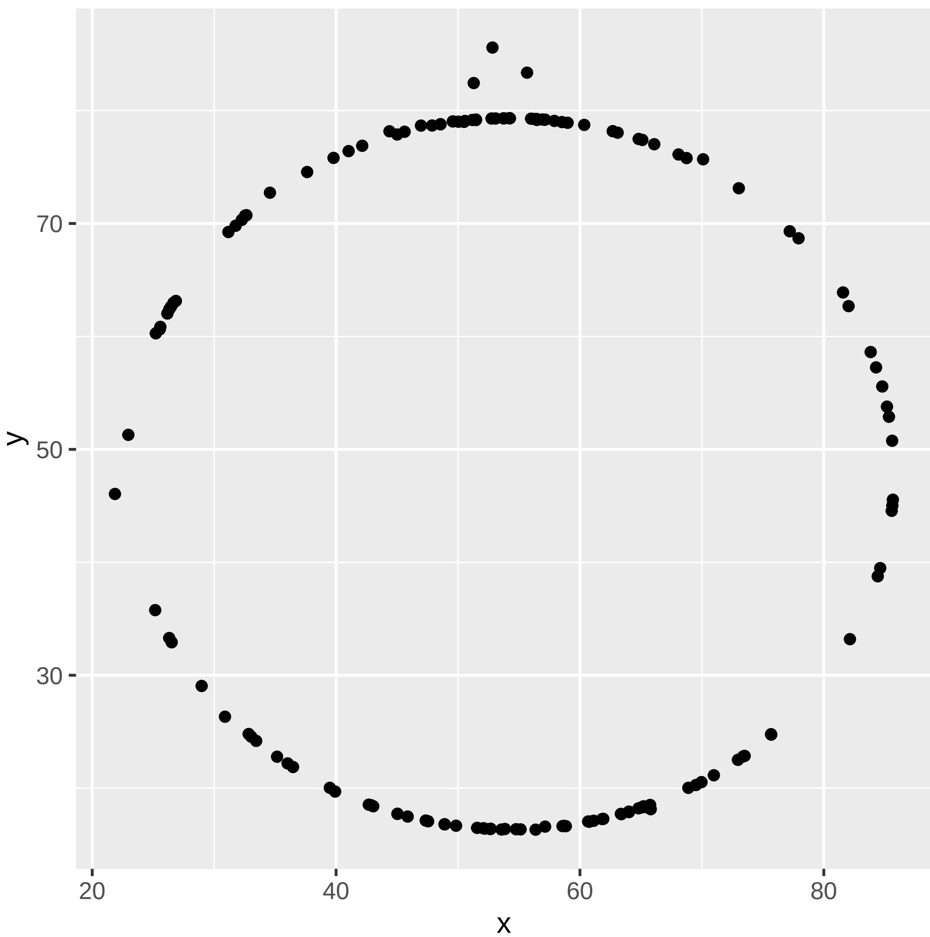
```
## # A tibble: 1 x 1
##         r
##     <dbl>
## 1 -0.0630
```

Exercise 4

(Add code and narrative as needed. Note that two `R` chunks are given but they are not labeled. Use the convention from Exercise 3 to name them appropriately)

```
circle_data <- datasaurus_dozen %>%
  filter(dataset == "circle")

ggplot(data = circle_data, mapping = aes(x = x, y = y)) +
  geom_point()
```

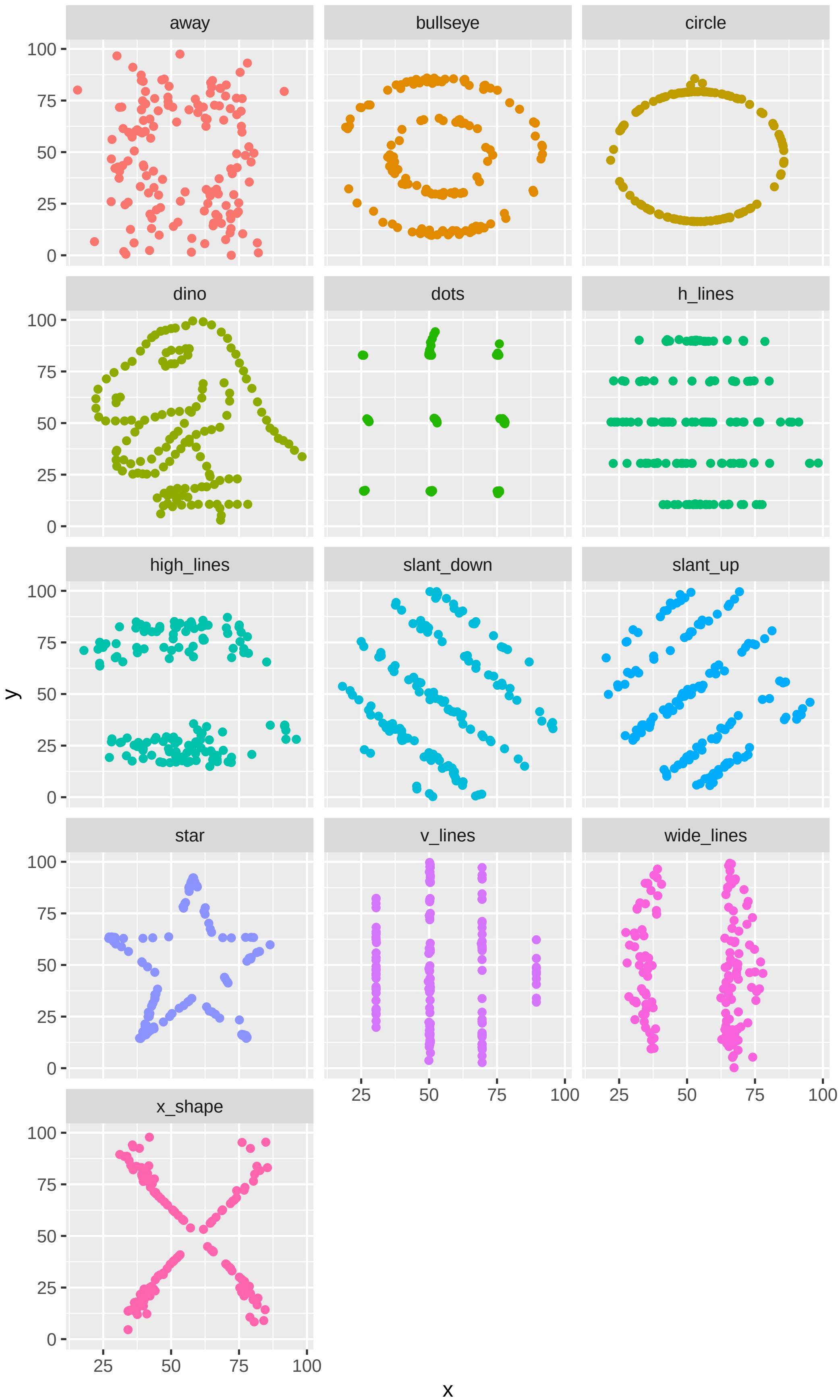



```
circle_data %>%
  summarize(r = cor(x, y))
```

```
## # A tibble: 1 x 1
##          r
##      <dbl>
## 1 -0.0683
```

Exercise 5

```
ggplot(datasaurus_dozen, aes(x = x, y = y, color = dataset))+
  geom_point()+
  facet_wrap(~ dataset, ncol = 3) +
  theme(legend.position = "none")
```


away
bullseye
circle
dino
dots
h_lines
high_lines
slant_down
slant_up
star
v_lines
wide_lines
x_shape
y
x
0
25
50
75
100

```r
datasaurus_dozen %>%
  group_by(dataset) %>%
  summarize(r = cor(x, y))
```

```
## # A tibble: 13 x 2
##    dataset          r
##    <chr>        <dbl>
##  1 away       -0.0641
##  2 bullseye   -0.0686
##  3 circle     -0.0683
##  4 dino       -0.0645
##  5 dots       -0.0603
##  6 h_lines    -0.0617
##  7 high_lines -0.0685
##  8 slant_down -0.0690
##  9 slant_up   -0.0686
## 10 star       -0.0630
## 11 v_lines    -0.0694
## 12 wide_lines -0.0666
## 13 x_shape    -0.0656
```

All of the correlation coefficients are relatively similar. However, those that are more negative have a stronger linear relationship compared to those that are closer to 0.