Lab 01 - Hello R

Harrison Kane

12/26/2022

Load packages

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

Exercise 1

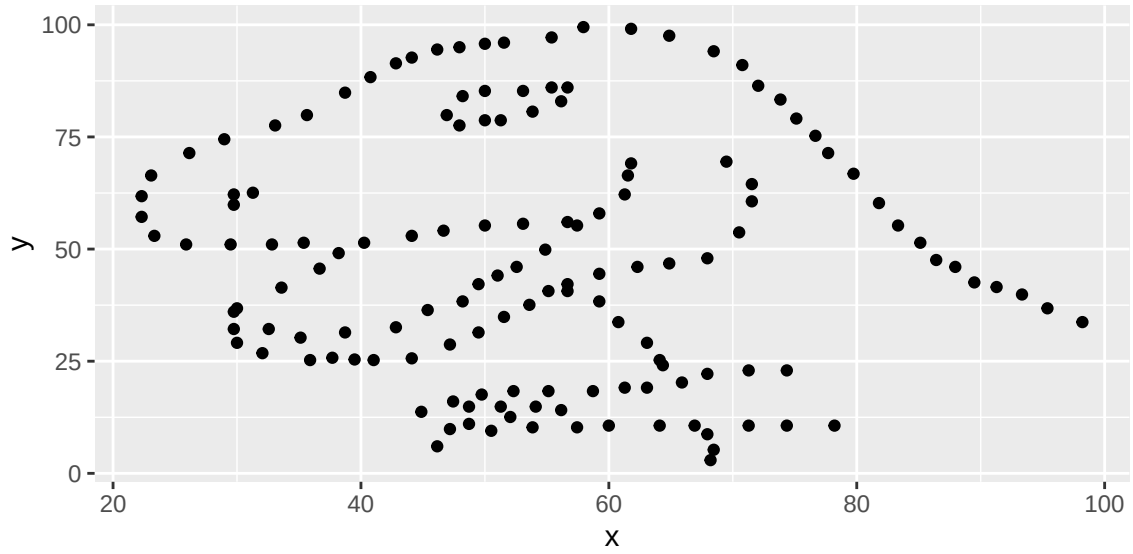
1846 rows, 3 columns

Exercise 2

First let's plot the data in the dino dataset:

```
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:

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

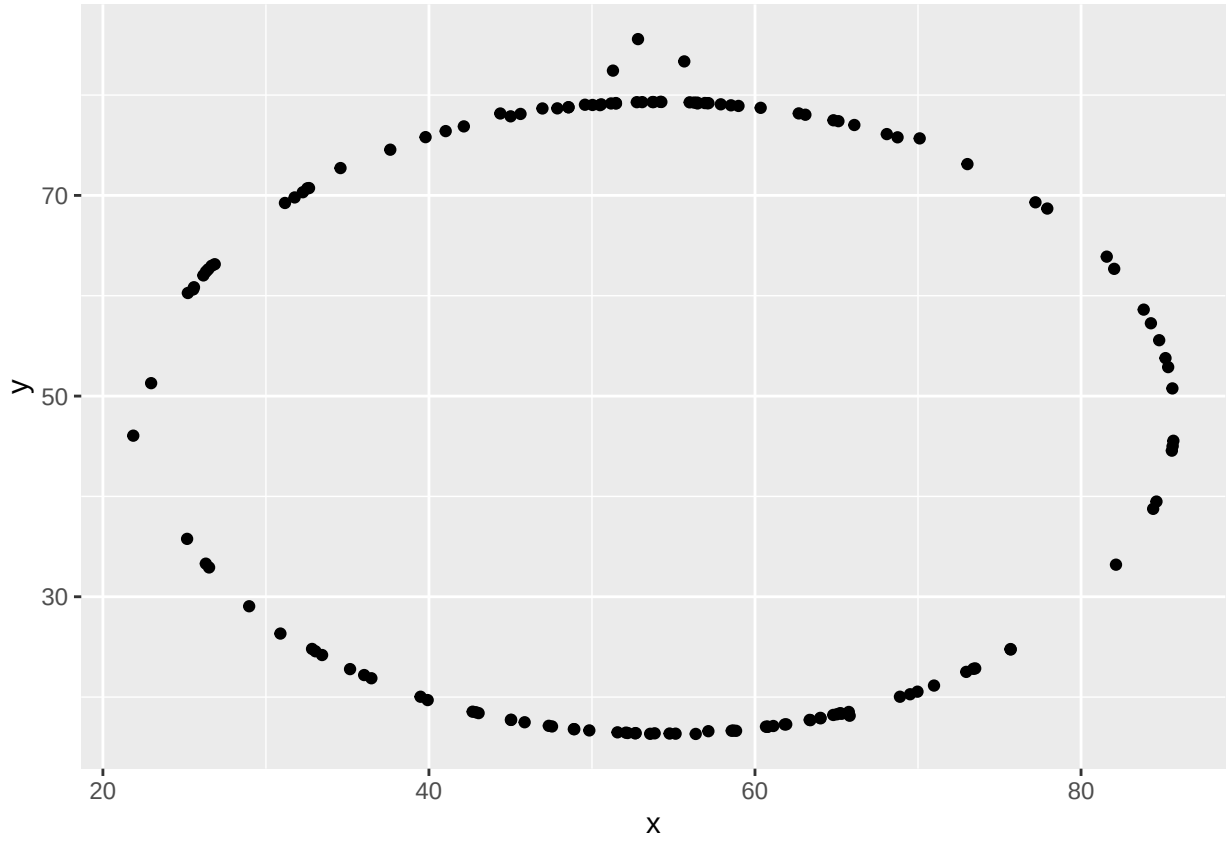
```
## # A tibble: 1 x 1
```

```
##         r
##     <dbl>
## 1 -0.0645
```

Exercise 3

First let's plot the data in the circle dataset:

```
circle_data <- datasaurus_dozen |>
    filter(dataset == "circle")
ggplot(data=circle_data, mapping = aes(x=x, y=y)) + geom_point()
```



And next calculate the correlation between x and y in this dataset:

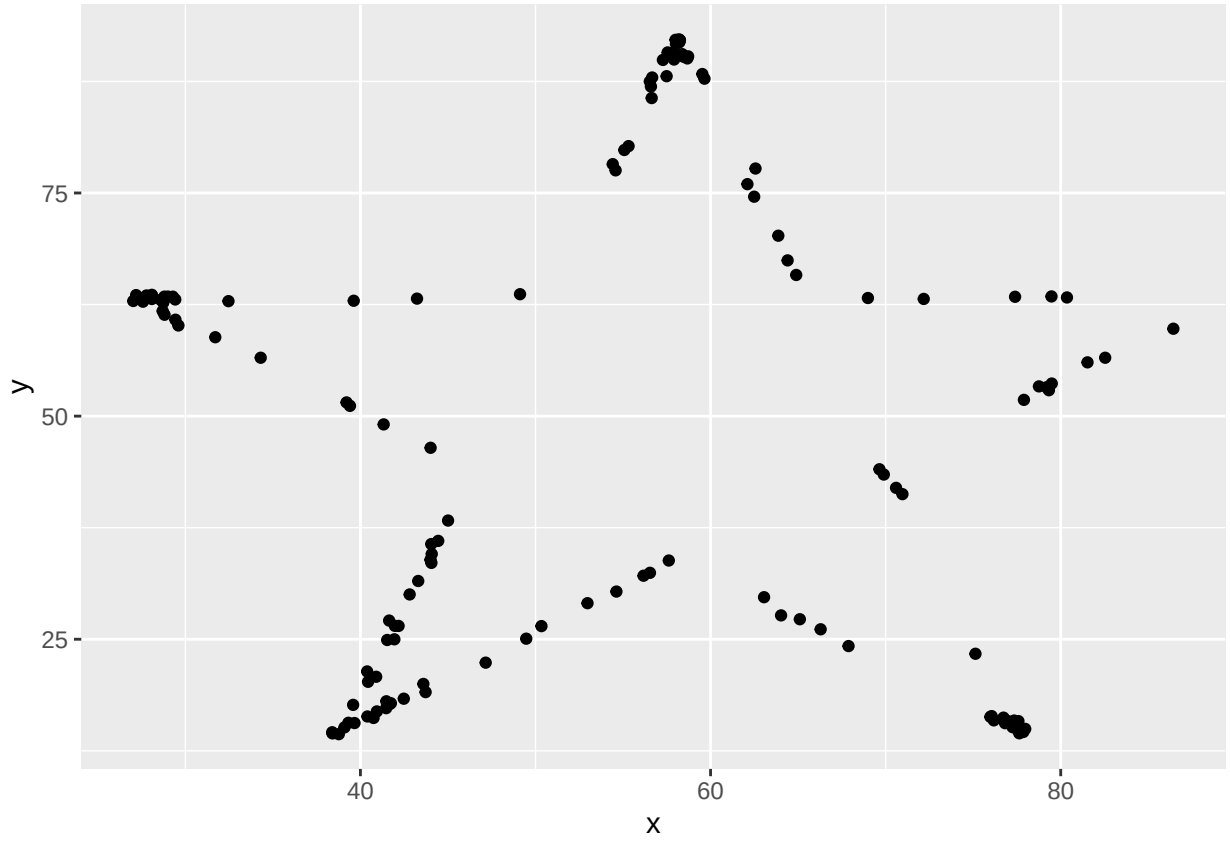
```
circle_data |>
  summarize(r = cor(x,y))
```

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

Exercise 4

First let's plot the data in the star dataset:

```
star_data <- datasaurus_dozen |>
    filter(dataset == "star")
ggplot(data=star_data, mapping = aes(x=x, y=y)) + geom_point()
```



And next calculate the correlation between x and y in this dataset:

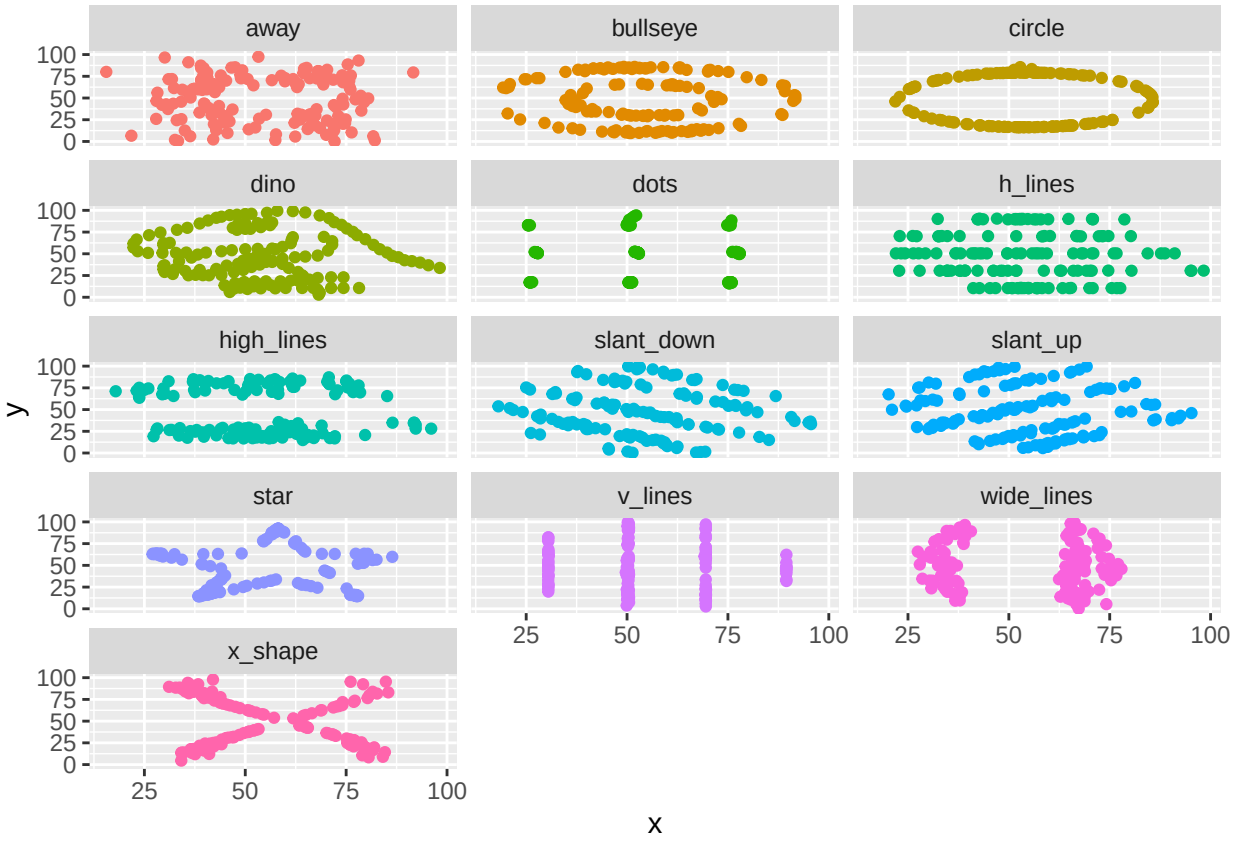
```
star_data |>
  summarize(r = cor(x,y))
```

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

Exercise 5

First let's plot all 13 datasets in the datasaurus_dozen:

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



Now calculate all the correlation coefficients:

```
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 datasets in the datasaurus_dozen have a very slightly negative correlation between x and y (roughly r = -0.06 for all 13 datasets), indicating little relationship between the variables.