

# Lab 01 - Hello R

---

Aleksandra Medina October 1, 2020

## Load packages

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

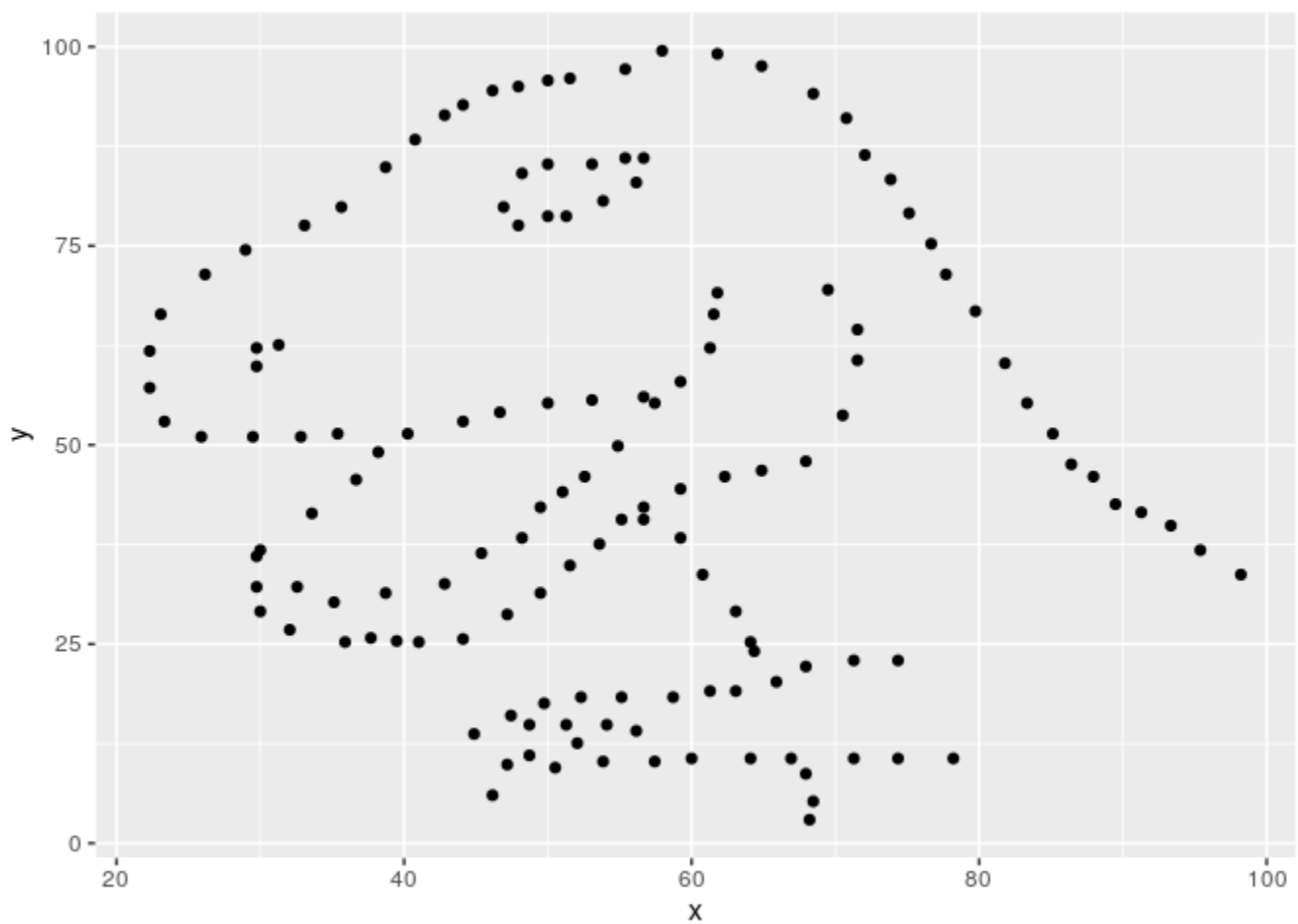
## Exercise 1

The `datasaurus_dozen` file has 1846 rows and 3 columns. The variables included are `dataset` (indicates which dataset the data are from), `x` (x-values), and `y` (y-values).

## Exercise 2

Plotting the dinosaur:

```
dino_data <- datasaurus_dozen %>%
  filter(dataset == "dino")
ggplot(data = dino_data, mapping = aes(x = x, y = y)) +
  geom_point()
```



The correlation coefficient between x and y for this dataset  $r = -0.0645$ .

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

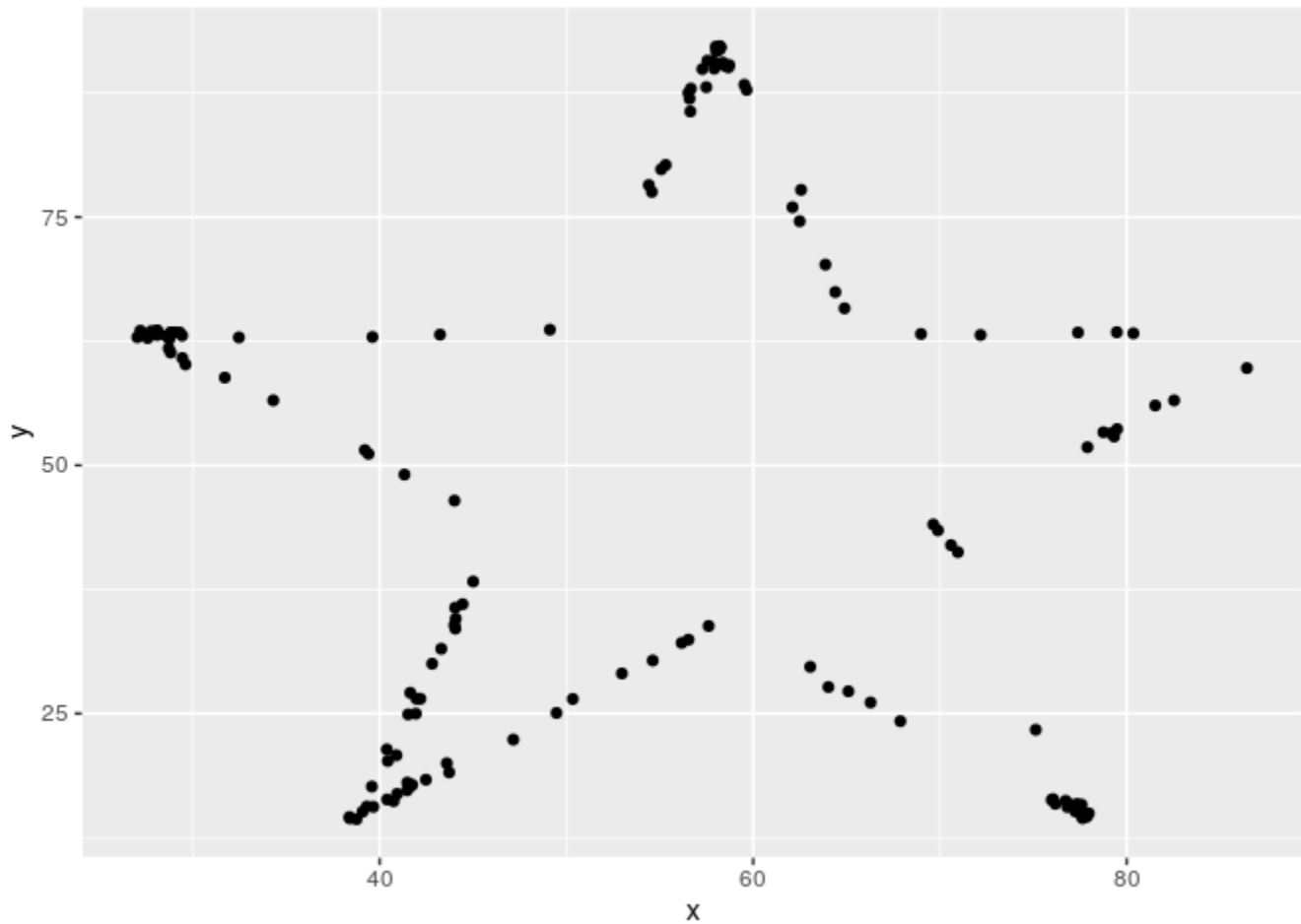
```
## # A tibble: 1 x 1
##       r
##   <dbl>
## 1 -0.0645
```

## Exercise 3

Plotting the star:

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

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



The correlation coefficient between x and y for this dataset  $r = -0.0630$ , compared to dino dataset  $r = -0.0645$ .

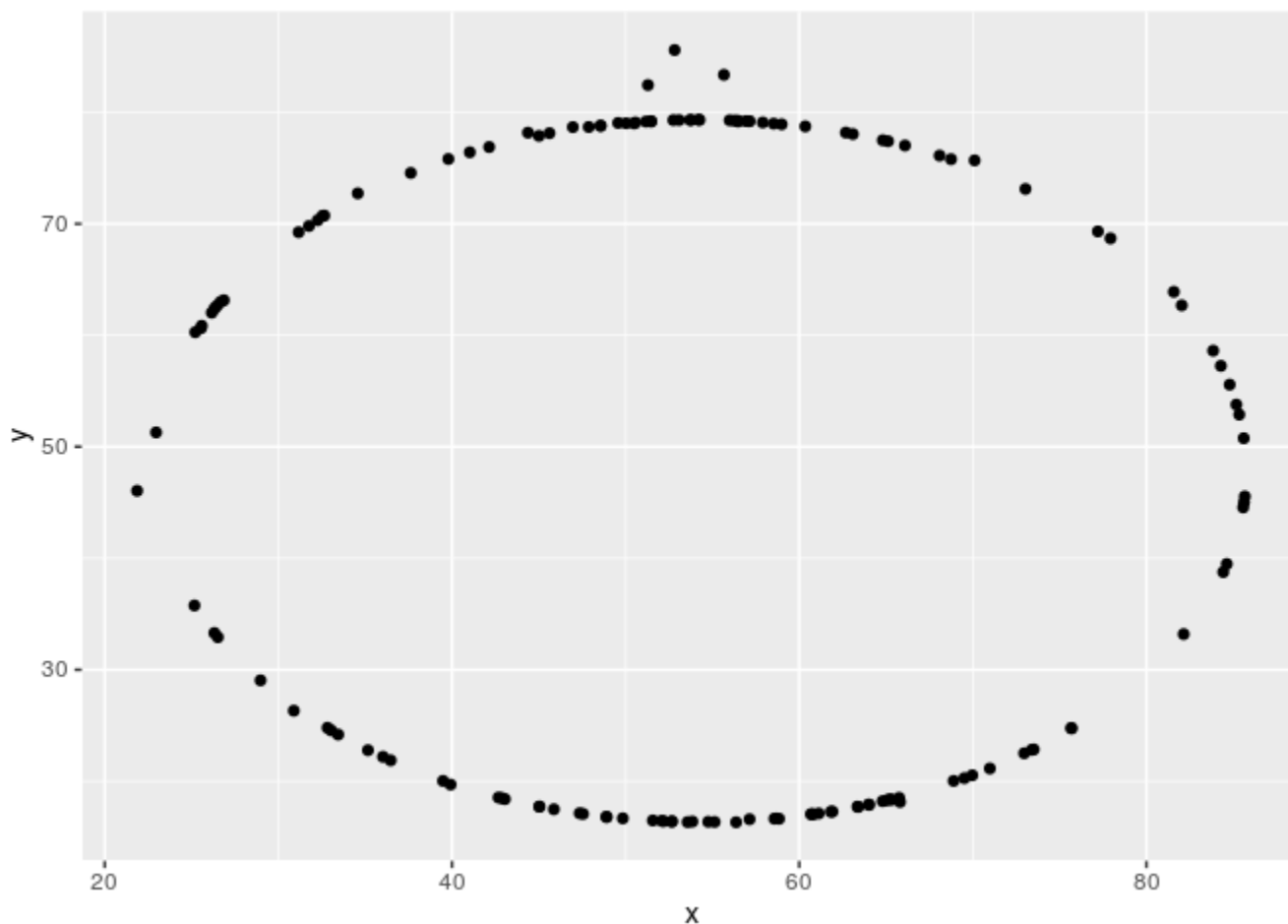
```
star_data %>%
  summarize(r = cor(x, y))
```

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

## Exercise 4

Plotting the circle:

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



The correlation coefficient between x and y for this dataset  $r = -0.0683$ , compared to dino dataset  $r = -0.0645$ .

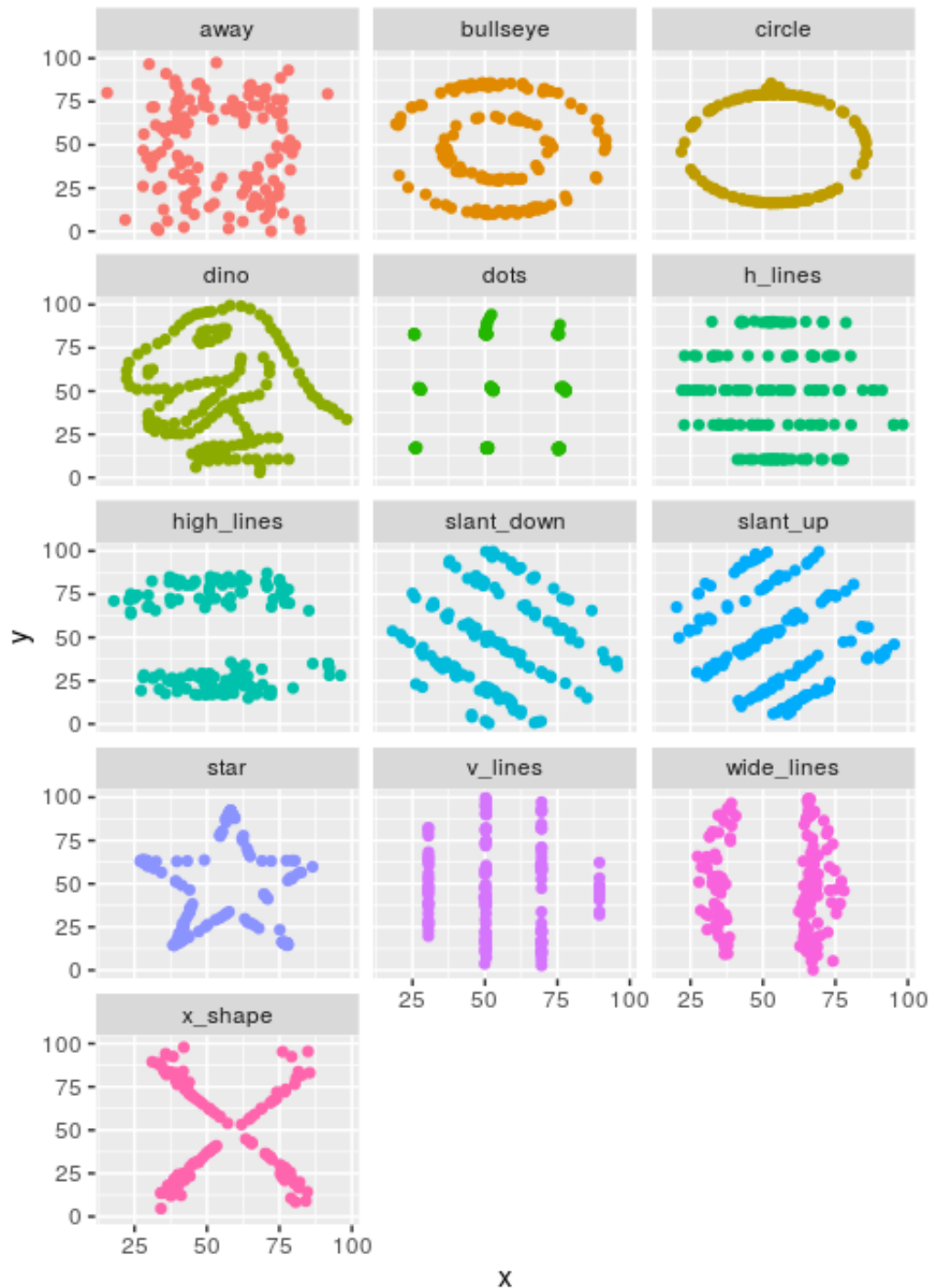
```
circle_data %>%
  summarize(r = cor(x, y))
```

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

## Exercise 5

Plotting all of the graphs together:

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



Calculating all of the r values at once:

- away:  $r=-0.0641$
- bullseye:  $r=-0.0686$
- circle:  $r=-0.0683$
- dino:  $r=-0.0645$
- dots:  $r=-0.0603$
- h\_lines:  $r=-0.0617$
- high\_lines:  $r=-0.0617$
- slant\_down:  $r=-0.0690$
- slant\_up:  $r=-0.0686$
- star:  $r=-0.0630$
- v\_lines:  $r=0.0694$
- wide\_lines:  $r=0.0666$
- x\_shape:  $r=-0.0656$

```
datasaurus_dozen %>%  
  group_by(dataset) %>%  
  summarize(r = cor(x, y)) %>%  
  print(13)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # 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
```

