Practical 0

K Pradeep Reddy

2025-05-16

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.

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. You can embed an R code chunk like this:

```r
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:
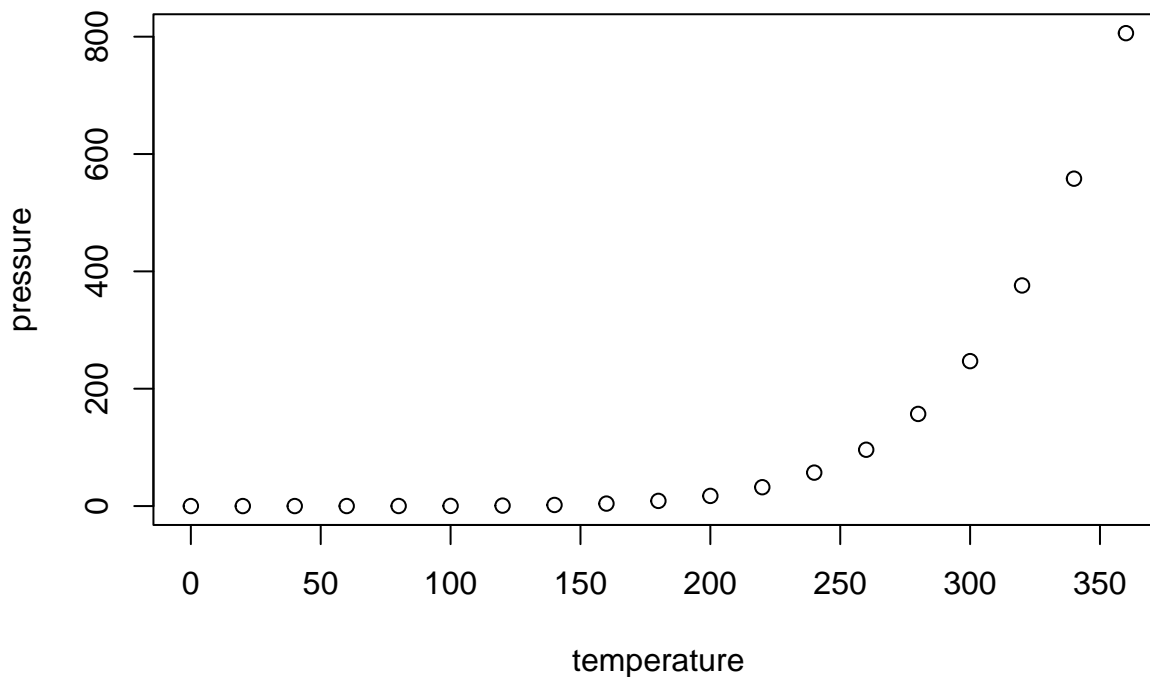
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.2      v tibble    3.2.1
## v lubridate 1.9.4      v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(datasauRus)

# Check structure
str(datasaurus_dozen)
```

```
## tibble [1,846 x 3] (S3: tbl_df/tbl/data.frame)
##  $ dataset: chr [1:1846] "dino" "dino" "dino" "dino" ...
##  $ x      : num [1:1846] 55.4 51.5 46.2 42.8 40.8 ...
##  $ y      : num [1:1846] 97.2 96 94.5 91.4 88.3 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   dataset = col_character(),
##   ..   x = col_double(),
##   ..   y = col_double()
##   .. )
```

```r
# Answer (write this as text below the chunk):
# The datasaurus_dozen dataset has 1846 rows and 3 columns: dataset, x, y
```

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

ggplot(data = dino_data, aes(x = x, y = y)) +
  geom_point()
```
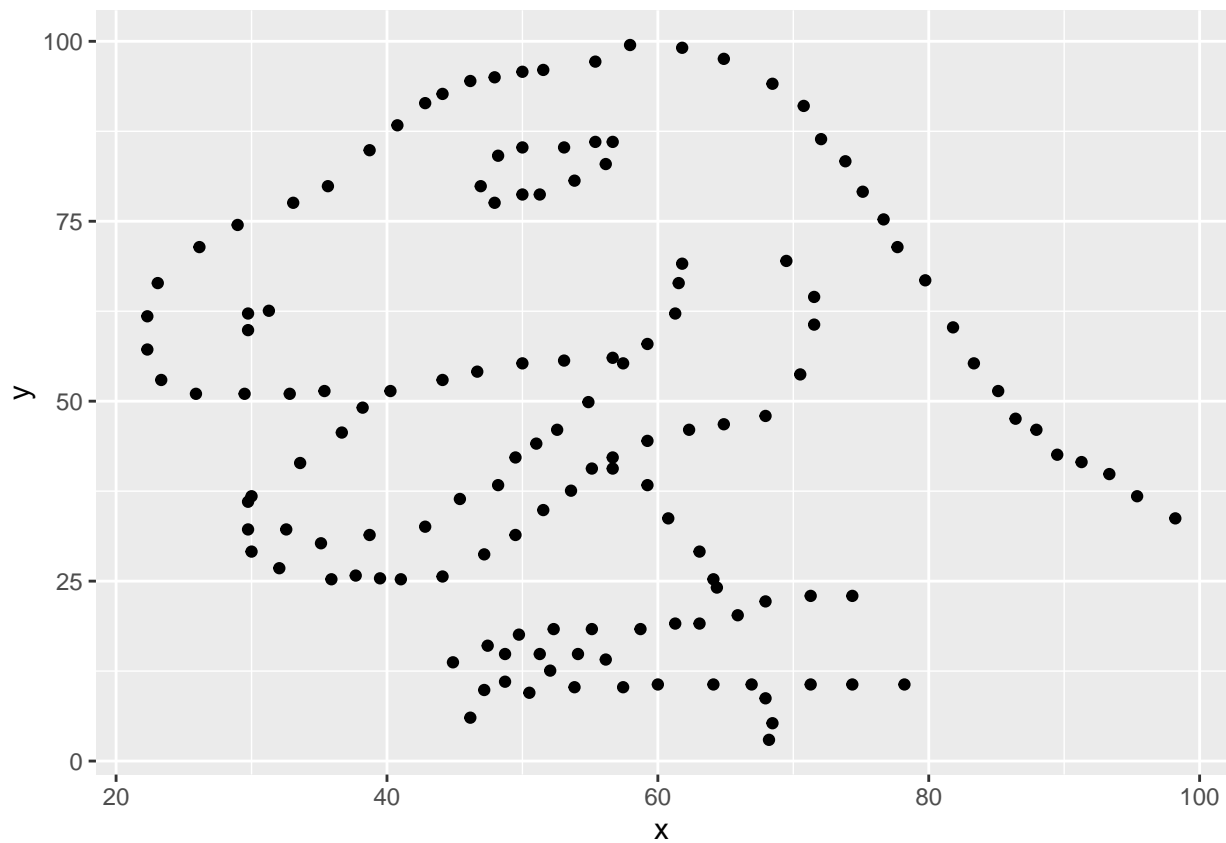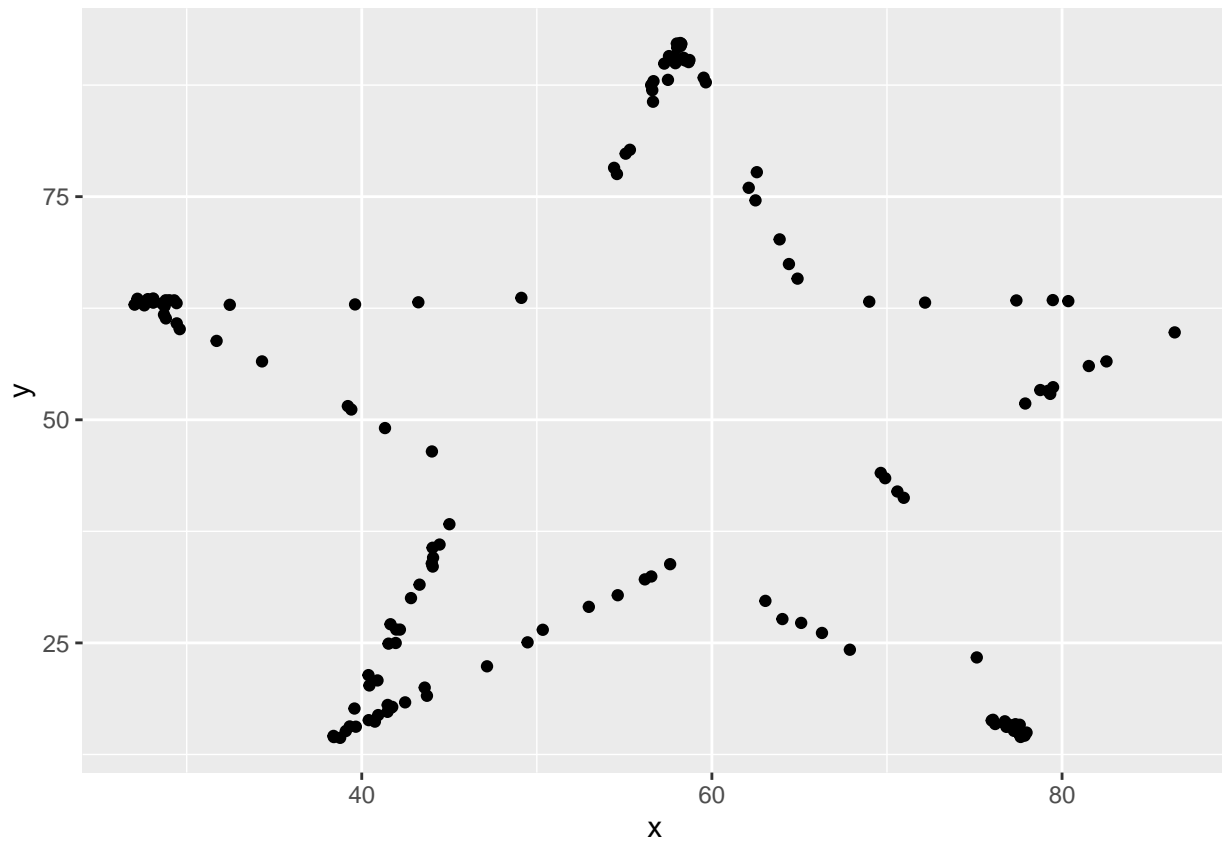
```
dino_data %>%
  summarize(r = cor(x, y))
```

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

```
# Example text: The correlation coefficient (r) between x and y for the dino dataset is approximately -0
```

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

ggplot(data = star_data, aes(x = x, y = y)) +
  geom_point()
```
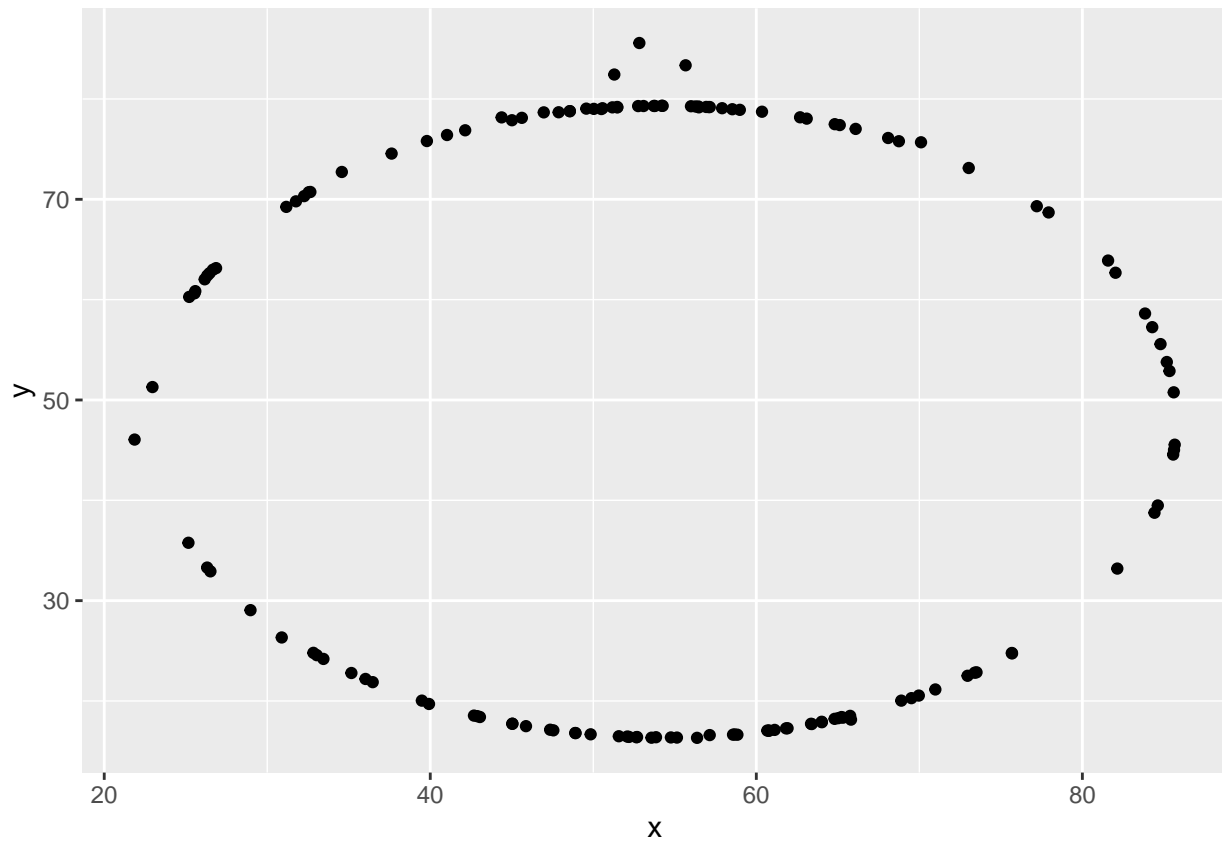
```
star_data %>%
  summarize(r = cor(x, y))
```

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

```
# Compare r to dino in your text answer
```

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

ggplot(data = circle_data, aes(x = x, y = y)) +
  geom_point()
```
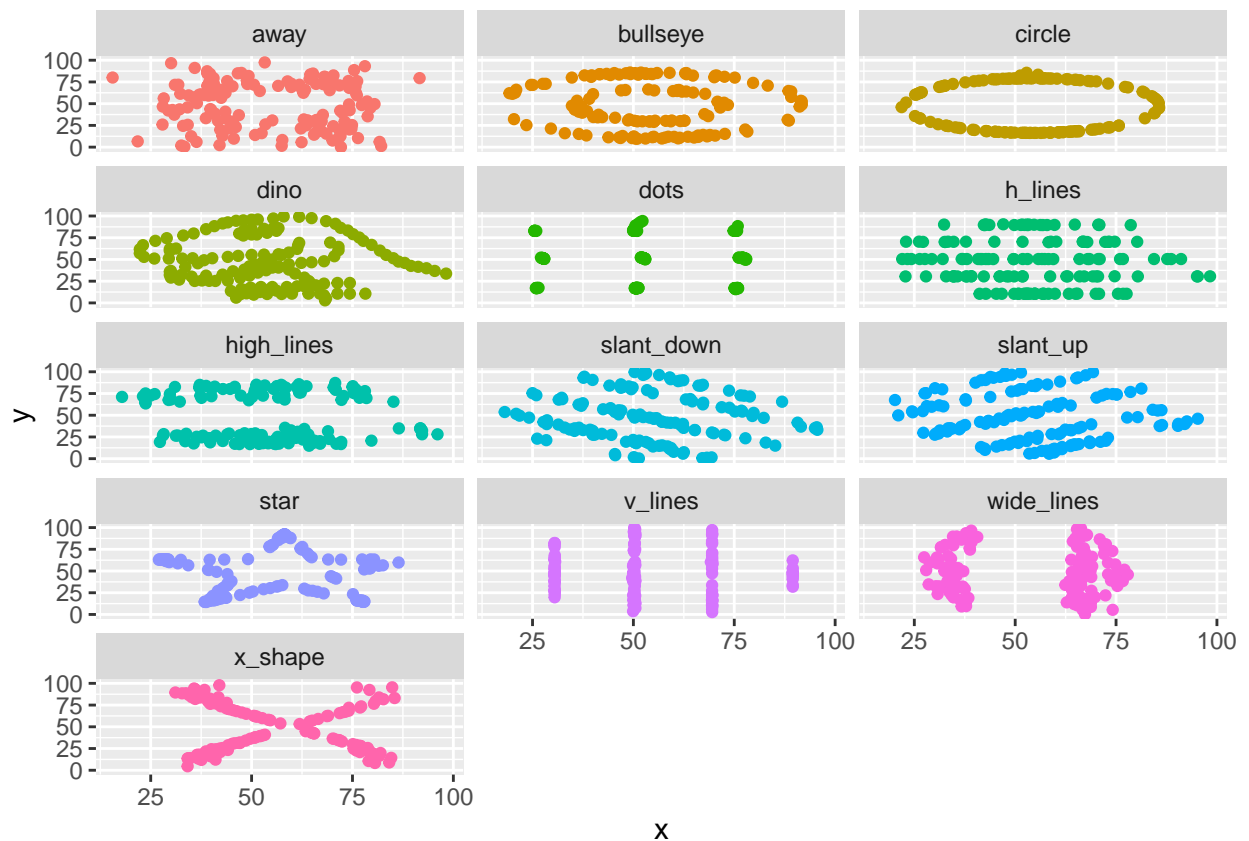
```
circle_data %>%
  summarize(r = cor(x, y))
```

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

```
# Compare and interpret in your text
```

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

```
# Correlation values for each dataset
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
```

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