R Markdown Template

Max Shteyman & Dr Qiu

2/10/2022

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 R codes and figures.

```
library(readr)
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

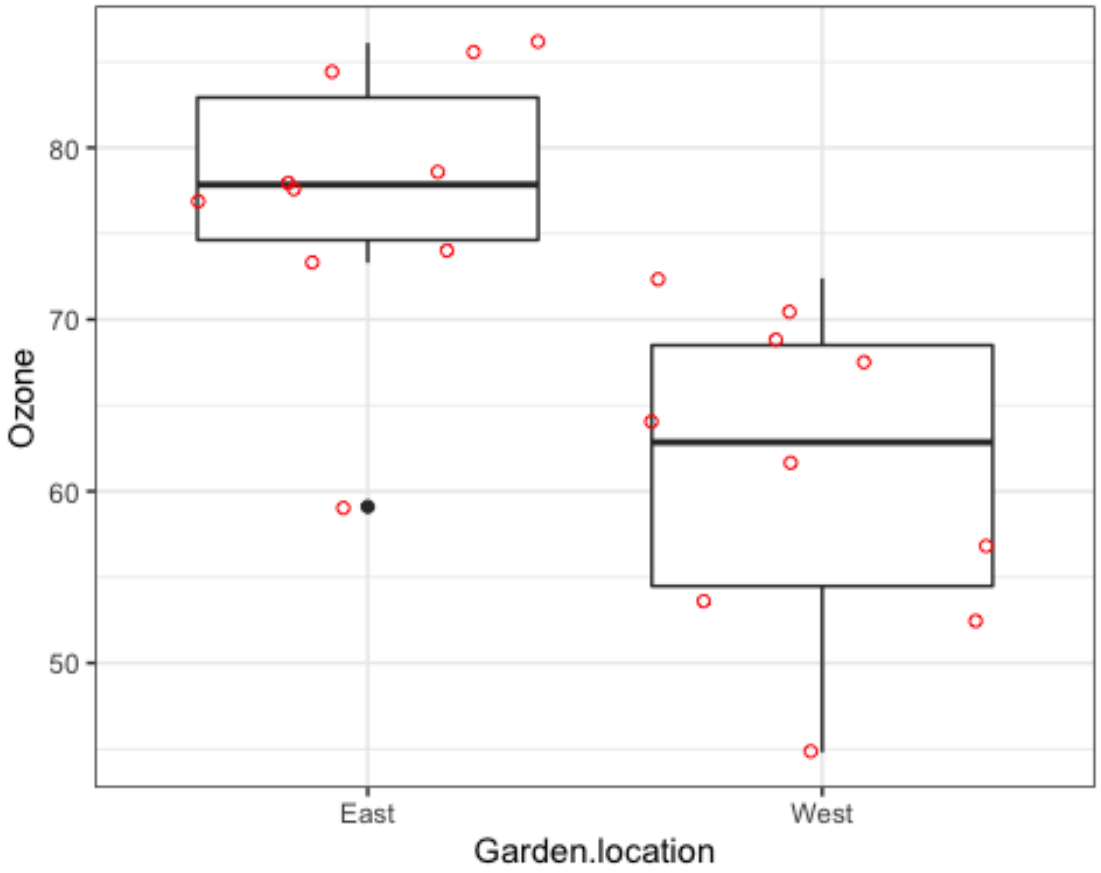
Read the ozone data set

```
ozone <- read_csv("ozone.csv")

## Rows: 20 Columns: 3
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (2): Garden.location, Garden.ID
## dbl (1): Ozone
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

glimpse(ozone) # returns a transposed version of the data frame, up to some
display limit. Also shows number of rows and columns and variable types.
```

```
## Rows: 20
## Columns: 3
## $ Ozone           <dbl> 61.7, 64.0, 72.4, 56.8, 52.4, 44.8, 70.4, 67.6,
68.8, …
## $ Garden.location <chr> "West", "West", "West", "West", "West", "West",
"West"…
## $ Garden.ID       <chr> "G1", "G2", "G3", "G4", "G5", "G6", "G7", "G8",
"G9", …
```

#two sample t-test data visualization boxplot

```
ggplot(data = ozone, aes(x = Garden.location, y =
Ozone)) + geom_boxplot() +
geom_jitter(shape=1,
color="red") + theme_bw()
```



#two sample test run t-test do a t.test now…

```
t.test(Ozone ~ Garden.location, data = ozone)

##
##  Welch Two Sample t-test
##
## data:  Ozone by Garden.location
```

```
## t = 4.2363, df = 17.656, p-value = 0.0005159
## alternative hypothesis: true difference in means between group East and
group West is not equal to 0
## 95 percent confidence interval:
##   8.094171 24.065829
## sample estimates:
## mean in group East mean in group West
##              77.34              61.26
```

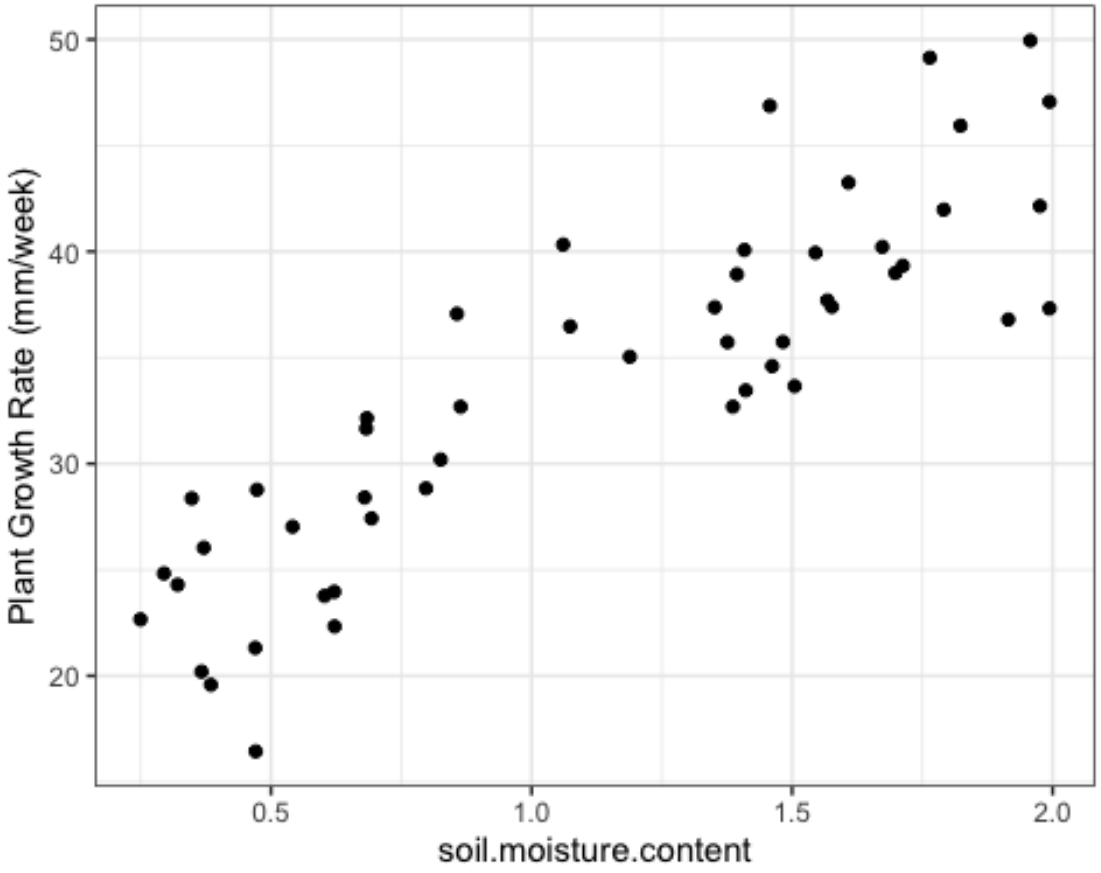
#Linear Regression Data & Question #Biological hypothesis: Soil moisture affect growth rate Statistical Hypothesis (H0): No correlation (r=0)

```
plant_gr <- read_csv("plant.growth.rate.csv")

## Rows: 50 Columns: 2
## ── Column specification
──────────────────────────────────────────────────────
## Delimiter: ","
## dbl (2): soil.moisture.content, plant.growth.rate
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

cols(
  soil.moisture.content = col_double () ,
  plant.growth.rate = col_double ())

## cols(
##   soil.moisture.content = col_double(),
##   plant.growth.rate = col_double()
## )

tbl_df(plant_gr)

## Warning: `tbl_df()` was deprecated in dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.

## # A tibble: 50 × 2
##    soil.moisture.content plant.growth.rate
##                    <dbl>             <dbl>
##  1                 0.470              21.3
##  2                 0.541              27.0
##  3                 1.70               39.0
##  4                 0.826              30.2
##  5                 0.857              37.1
##  6                 1.61               43.2
##  7                 0.250              22.7
```

```
##  8                 1.67               40.2
##  9                 1.46               46.9
## 10                 0.473              28.8
## # … with 40 more rows
```

#Linear Regression Visualization

```
ggplot(plant_gr,
       aes(x = soil.moisture.content, y = plant.growth.rate)) +
          geom_point () +
          ylab("Plant Growth Rate (mm/week)") +
          theme_bw()
```



#Linear Regression Run linear model

```
model_pgr <- lm(plant.growth.rate ~ soil.moisture.content,
                data = plant_gr)
summary(model_pgr)

##
## Call:
## lm(formula = plant.growth.rate ~ soil.moisture.content, data = plant_gr)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -8.9089 -3.0747  0.2261  2.6567  8.9406
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)              19.348      1.283   15.08   <2e-16 ***
## soil.moisture.content    12.750      1.021   12.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.019 on 48 degrees of freedom
## Multiple R-squared:  0.7648, Adjusted R-squared:  0.7599
## F-statistic: 156.1 on 1 and 48 DF,  p-value: < 2.2e-16
```

#Linear Regression Re-plot (add regression line & confidence band)

```
ggplot(plant_gr, aes(x = soil.moisture.content,
          y = plant.growth.rate)) +
  geom_point() +
  geom_smooth(method = 'lm') +
  ylab("Plant Growth Rate (mm/week)") +
  theme_bw()
```

```
## `geom_smooth()` using formula 'y ~ x'
```