Week3

your name, D Toher, B Derrick

insert date

# installing packages used in the exercise today

## Loading required package: ggplot2

## Loading required package: flextable

## Loading required package: psych

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

## Loading required package: tableone

# Introduction to Fisher’s Iris dataset

This example is famous and you will find a lot of information about it particularly in R user guides.

The dataset is within R so you do not to read in a data file.

A good starting point to understand the data is to look at the data structure (using the function str) and look at some basic descriptive statistics of each of the variables (using the function summary)

str(iris)

## 'data.frame': 150 obs. of 5 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

summary(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500   
## Species   
## setosa :50   
## versicolor:50   
## virginica :50   
##   
##   
##

summary(iris$Sepal.Length)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.300 5.100 5.800 5.843 6.400 7.900

summary(iris$Sepal.Width)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 2.800 3.000 3.057 3.300 4.400

summary(iris$Petal.Length)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.600 4.350 3.758 5.100 6.900

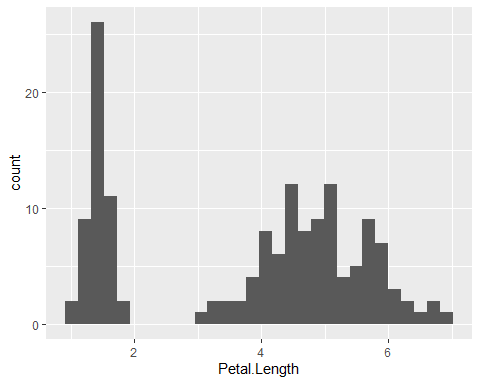
summary(iris$Petal.Width)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.100 0.300 1.300 1.199 1.800 2.500

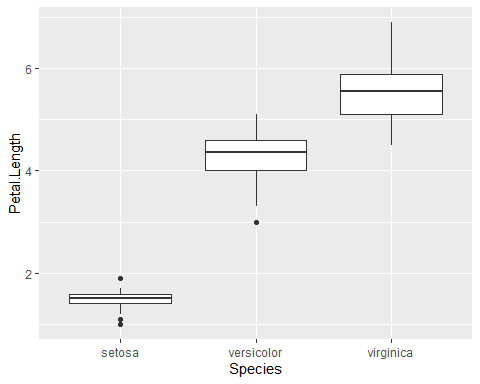
## Example graphics

ggplot(data=iris)+  
 geom\_histogram(mapping=aes(x=Petal.Length))

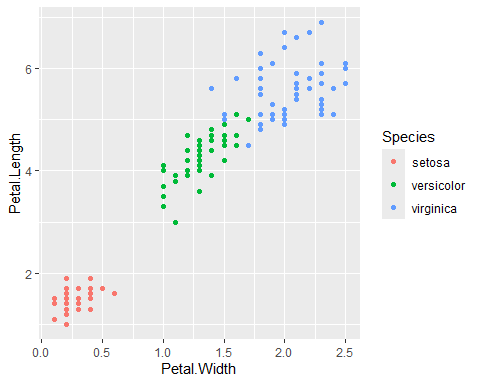
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggplot(data=iris,aes(x=Species,y=Petal.Length))+  
 geom\_boxplot()



ggplot(data=iris,aes(x=Petal.Width,y=Petal.Length, color=Species))+  
 geom\_point()



Discussion:

1. what do the above graphics show?
2. which of the above graphics are most appropriate?
3. what additional graphics may be useful?

# Dewis exercise 1: R\_filtering\_sampling.

Sometimes you may be asked to work with a subset of data, either excluding a particular group, or taking a sample from a larger dataset.

Insert your code to complete the Dewis exercise here:

## Additional graphics for the iris datsest

Attempt some additional relevant graphics using the iris dataset below:

Interpret your graphics

# Dewis exercise 2: r\_descriptive\_statistics

In this exercise you will read in a dataset from Dewis using the function read.csv

You will then summarise the data and see if this reveals any issues.

Use this space to experiment with the R code provided to complete this task.

# Advanced: Nice looking tables using the additional packages installed

This shows how to start to get some really nice looking tables for the iris dataset

describeBy(iris[,-5],group=iris[,5])

##   
## Descriptive statistics by group   
## group: setosa  
## vars n mean sd median trimmed mad min max range skew kurtosis  
## Sepal.Length 1 50 5.01 0.35 5.0 5.00 0.30 4.3 5.8 1.5 0.11 -0.45  
## Sepal.Width 2 50 3.43 0.38 3.4 3.42 0.37 2.3 4.4 2.1 0.04 0.60  
## Petal.Length 3 50 1.46 0.17 1.5 1.46 0.15 1.0 1.9 0.9 0.10 0.65  
## Petal.Width 4 50 0.25 0.11 0.2 0.24 0.00 0.1 0.6 0.5 1.18 1.26  
## se  
## Sepal.Length 0.05  
## Sepal.Width 0.05  
## Petal.Length 0.02  
## Petal.Width 0.01  
## ------------------------------------------------------------   
## group: versicolor  
## vars n mean sd median trimmed mad min max range skew kurtosis  
## Sepal.Length 1 50 5.94 0.52 5.90 5.94 0.52 4.9 7.0 2.1 0.10 -0.69  
## Sepal.Width 2 50 2.77 0.31 2.80 2.78 0.30 2.0 3.4 1.4 -0.34 -0.55  
## Petal.Length 3 50 4.26 0.47 4.35 4.29 0.52 3.0 5.1 2.1 -0.57 -0.19  
## Petal.Width 4 50 1.33 0.20 1.30 1.32 0.22 1.0 1.8 0.8 -0.03 -0.59  
## se  
## Sepal.Length 0.07  
## Sepal.Width 0.04  
## Petal.Length 0.07  
## Petal.Width 0.03  
## ------------------------------------------------------------   
## group: virginica  
## vars n mean sd median trimmed mad min max range skew kurtosis  
## Sepal.Length 1 50 6.59 0.64 6.50 6.57 0.59 4.9 7.9 3.0 0.11 -0.20  
## Sepal.Width 2 50 2.97 0.32 3.00 2.96 0.30 2.2 3.8 1.6 0.34 0.38  
## Petal.Length 3 50 5.55 0.55 5.55 5.51 0.67 4.5 6.9 2.4 0.52 -0.37  
## Petal.Width 4 50 2.03 0.27 2.00 2.03 0.30 1.4 2.5 1.1 -0.12 -0.75  
## se  
## Sepal.Length 0.09  
## Sepal.Width 0.05  
## Petal.Length 0.08  
## Petal.Width 0.04

describeBy(iris[,-5],group=iris[,5],skew=FALSE)

##   
## Descriptive statistics by group   
## group: setosa  
## vars n mean sd median min max range se  
## Sepal.Length 1 50 5.01 0.35 5.0 4.3 5.8 1.5 0.05  
## Sepal.Width 2 50 3.43 0.38 3.4 2.3 4.4 2.1 0.05  
## Petal.Length 3 50 1.46 0.17 1.5 1.0 1.9 0.9 0.02  
## Petal.Width 4 50 0.25 0.11 0.2 0.1 0.6 0.5 0.01  
## ------------------------------------------------------------   
## group: versicolor  
## vars n mean sd median min max range se  
## Sepal.Length 1 50 5.94 0.52 5.90 4.9 7.0 2.1 0.07  
## Sepal.Width 2 50 2.77 0.31 2.80 2.0 3.4 1.4 0.04  
## Petal.Length 3 50 4.26 0.47 4.35 3.0 5.1 2.1 0.07  
## Petal.Width 4 50 1.33 0.20 1.30 1.0 1.8 0.8 0.03  
## ------------------------------------------------------------   
## group: virginica  
## vars n mean sd median min max range se  
## Sepal.Length 1 50 6.59 0.64 6.50 4.9 7.9 3.0 0.09  
## Sepal.Width 2 50 2.97 0.32 3.00 2.2 3.8 1.6 0.05  
## Petal.Length 3 50 5.55 0.55 5.55 4.5 6.9 2.4 0.08  
## Petal.Width 4 50 2.03 0.27 2.00 1.4 2.5 1.1 0.04

IrisSummary<-describeBy(iris[,-5],group=iris[,5],skew=FALSE)  
IrisSummary2<-describeBy(iris[,-5],group=iris[,5],skew=FALSE,mat=TRUE)  
#View(IrisSummary2)  
dim(IrisSummary)

## [1] 3

#colnames(iris)[-5]  
  
# which columns are item and vars?  
Ignore<-colnames(IrisSummary2)%in% c("item","vars")  
  
#Ignore  
IrisSum<-IrisSummary2[,!Ignore]  
  
#colnames(IrisSum)  
colnames(IrisSum)[colnames(IrisSum)=="group1"]<-"species"  
  
  
  
  
flextable(IrisSum)

| species | n | mean | sd | median | min | max | range | se |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| setosa | 50 | 5.006 | 0.3524897 | 5.00 | 4.3 | 5.8 | 1.5 | 0.04984957 |
| versicolor | 50 | 5.936 | 0.5161711 | 5.90 | 4.9 | 7.0 | 2.1 | 0.07299762 |
| virginica | 50 | 6.588 | 0.6358796 | 6.50 | 4.9 | 7.9 | 3.0 | 0.08992695 |
| setosa | 50 | 3.428 | 0.3790644 | 3.40 | 2.3 | 4.4 | 2.1 | 0.05360780 |
| versicolor | 50 | 2.770 | 0.3137983 | 2.80 | 2.0 | 3.4 | 1.4 | 0.04437778 |
| virginica | 50 | 2.974 | 0.3224966 | 3.00 | 2.2 | 3.8 | 1.6 | 0.04560791 |
| setosa | 50 | 1.462 | 0.1736640 | 1.50 | 1.0 | 1.9 | 0.9 | 0.02455980 |
| versicolor | 50 | 4.260 | 0.4699110 | 4.35 | 3.0 | 5.1 | 2.1 | 0.06645545 |
| virginica | 50 | 5.552 | 0.5518947 | 5.55 | 4.5 | 6.9 | 2.4 | 0.07804970 |
| setosa | 50 | 0.246 | 0.1053856 | 0.20 | 0.1 | 0.6 | 0.5 | 0.01490377 |
| versicolor | 50 | 1.326 | 0.1977527 | 1.30 | 1.0 | 1.8 | 0.8 | 0.02796645 |
| virginica | 50 | 2.026 | 0.2746501 | 2.00 | 1.4 | 2.5 | 1.1 | 0.03884138 |

# need to add in a leading column with the variable information  
colnames(iris)[-5]

## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"

niceNames<-c("Sepal length","Sepal width","Petal length","Petal width")  
  
#rep(niceNames,each=3)  
IrisSum2<-cbind(variable=rep(niceNames,each=3),IrisSum)  
#IrisSum2  
  
flextable(IrisSum2)

| variable | species | n | mean | sd | median | min | max | range | se |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sepal length | setosa | 50 | 5.006 | 0.3524897 | 5.00 | 4.3 | 5.8 | 1.5 | 0.04984957 |
| Sepal length | versicolor | 50 | 5.936 | 0.5161711 | 5.90 | 4.9 | 7.0 | 2.1 | 0.07299762 |
| Sepal length | virginica | 50 | 6.588 | 0.6358796 | 6.50 | 4.9 | 7.9 | 3.0 | 0.08992695 |
| Sepal width | setosa | 50 | 3.428 | 0.3790644 | 3.40 | 2.3 | 4.4 | 2.1 | 0.05360780 |
| Sepal width | versicolor | 50 | 2.770 | 0.3137983 | 2.80 | 2.0 | 3.4 | 1.4 | 0.04437778 |
| Sepal width | virginica | 50 | 2.974 | 0.3224966 | 3.00 | 2.2 | 3.8 | 1.6 | 0.04560791 |
| Petal length | setosa | 50 | 1.462 | 0.1736640 | 1.50 | 1.0 | 1.9 | 0.9 | 0.02455980 |
| Petal length | versicolor | 50 | 4.260 | 0.4699110 | 4.35 | 3.0 | 5.1 | 2.1 | 0.06645545 |
| Petal length | virginica | 50 | 5.552 | 0.5518947 | 5.55 | 4.5 | 6.9 | 2.4 | 0.07804970 |
| Petal width | setosa | 50 | 0.246 | 0.1053856 | 0.20 | 0.1 | 0.6 | 0.5 | 0.01490377 |
| Petal width | versicolor | 50 | 1.326 | 0.1977527 | 1.30 | 1.0 | 1.8 | 0.8 | 0.02796645 |
| Petal width | virginica | 50 | 2.026 | 0.2746501 | 2.00 | 1.4 | 2.5 | 1.1 | 0.03884138 |

merge\_v(flextable(IrisSum2))

| variable | species | n | mean | sd | median | min | max | range | se |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sepal length | setosa | 50 | 5.006 | 0.3524897 | 5.00 | 4.3 | 5.8 | 1.5 | 0.04984957 |
| versicolor | 5.936 | 0.5161711 | 5.90 | 4.9 | 7.0 | 2.1 | 0.07299762 |
| virginica | 6.588 | 0.6358796 | 6.50 | 7.9 | 3.0 | 0.08992695 |
| Sepal width | setosa | 3.428 | 0.3790644 | 3.40 | 2.3 | 4.4 | 2.1 | 0.05360780 |
| versicolor | 2.770 | 0.3137983 | 2.80 | 2.0 | 3.4 | 1.4 | 0.04437778 |
| virginica | 2.974 | 0.3224966 | 3.00 | 2.2 | 3.8 | 1.6 | 0.04560791 |
| Petal length | setosa | 1.462 | 0.1736640 | 1.50 | 1.0 | 1.9 | 0.9 | 0.02455980 |
| versicolor | 4.260 | 0.4699110 | 4.35 | 3.0 | 5.1 | 2.1 | 0.06645545 |
| virginica | 5.552 | 0.5518947 | 5.55 | 4.5 | 6.9 | 2.4 | 0.07804970 |
| Petal width | setosa | 0.246 | 0.1053856 | 0.20 | 0.1 | 0.6 | 0.5 | 0.01490377 |
| versicolor | 1.326 | 0.1977527 | 1.30 | 1.0 | 1.8 | 0.8 | 0.02796645 |
| virginica | 2.026 | 0.2746501 | 2.00 | 1.4 | 2.5 | 1.1 | 0.03884138 |

merge\_v(flextable(IrisSum2),j=1) # only merge 1st column!

| variable | species | n | mean | sd | median | min | max | range | se |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sepal length | setosa | 50 | 5.006 | 0.3524897 | 5.00 | 4.3 | 5.8 | 1.5 | 0.04984957 |
| versicolor | 50 | 5.936 | 0.5161711 | 5.90 | 4.9 | 7.0 | 2.1 | 0.07299762 |
| virginica | 50 | 6.588 | 0.6358796 | 6.50 | 4.9 | 7.9 | 3.0 | 0.08992695 |
| Sepal width | setosa | 50 | 3.428 | 0.3790644 | 3.40 | 2.3 | 4.4 | 2.1 | 0.05360780 |
| versicolor | 50 | 2.770 | 0.3137983 | 2.80 | 2.0 | 3.4 | 1.4 | 0.04437778 |
| virginica | 50 | 2.974 | 0.3224966 | 3.00 | 2.2 | 3.8 | 1.6 | 0.04560791 |
| Petal length | setosa | 50 | 1.462 | 0.1736640 | 1.50 | 1.0 | 1.9 | 0.9 | 0.02455980 |
| versicolor | 50 | 4.260 | 0.4699110 | 4.35 | 3.0 | 5.1 | 2.1 | 0.06645545 |
| virginica | 50 | 5.552 | 0.5518947 | 5.55 | 4.5 | 6.9 | 2.4 | 0.07804970 |
| Petal width | setosa | 50 | 0.246 | 0.1053856 | 0.20 | 0.1 | 0.6 | 0.5 | 0.01490377 |
| versicolor | 50 | 1.326 | 0.1977527 | 1.30 | 1.0 | 1.8 | 0.8 | 0.02796645 |
| virginica | 50 | 2.026 | 0.2746501 | 2.00 | 1.4 | 2.5 | 1.1 | 0.03884138 |

theme\_booktabs(merge\_v(flextable(IrisSum2),j=1))

| variable | species | n | mean | sd | median | min | max | range | se |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sepal length | setosa | 50 | 5.006 | 0.3524897 | 5.00 | 4.3 | 5.8 | 1.5 | 0.04984957 |
| versicolor | 50 | 5.936 | 0.5161711 | 5.90 | 4.9 | 7.0 | 2.1 | 0.07299762 |
| virginica | 50 | 6.588 | 0.6358796 | 6.50 | 4.9 | 7.9 | 3.0 | 0.08992695 |
| Sepal width | setosa | 50 | 3.428 | 0.3790644 | 3.40 | 2.3 | 4.4 | 2.1 | 0.05360780 |
| versicolor | 50 | 2.770 | 0.3137983 | 2.80 | 2.0 | 3.4 | 1.4 | 0.04437778 |
| virginica | 50 | 2.974 | 0.3224966 | 3.00 | 2.2 | 3.8 | 1.6 | 0.04560791 |
| Petal length | setosa | 50 | 1.462 | 0.1736640 | 1.50 | 1.0 | 1.9 | 0.9 | 0.02455980 |
| versicolor | 50 | 4.260 | 0.4699110 | 4.35 | 3.0 | 5.1 | 2.1 | 0.06645545 |
| virginica | 50 | 5.552 | 0.5518947 | 5.55 | 4.5 | 6.9 | 2.4 | 0.07804970 |
| Petal width | setosa | 50 | 0.246 | 0.1053856 | 0.20 | 0.1 | 0.6 | 0.5 | 0.01490377 |
| versicolor | 50 | 1.326 | 0.1977527 | 1.30 | 1.0 | 1.8 | 0.8 | 0.02796645 |
| virginica | 50 | 2.026 | 0.2746501 | 2.00 | 1.4 | 2.5 | 1.1 | 0.03884138 |

CreateTableOne(vars=names(iris)[-5],  
 strata="Species",   
 data=iris,  
 test=F)

## Stratified by Species  
## setosa versicolor virginica   
## n 50 50 50   
## Sepal.Length (mean (SD)) 5.01 (0.35) 5.94 (0.52) 6.59 (0.64)  
## Sepal.Width (mean (SD)) 3.43 (0.38) 2.77 (0.31) 2.97 (0.32)  
## Petal.Length (mean (SD)) 1.46 (0.17) 4.26 (0.47) 5.55 (0.55)  
## Petal.Width (mean (SD)) 0.25 (0.11) 1.33 (0.20) 2.03 (0.27)

# End matter - Session Information

It is usually good to make sure that you compile your document so that the information about your session (including software / package versions) is stated.

sessionInfo()

## R version 4.5.1 (2025-06-13 ucrt)  
## Platform: x86\_64-w64-mingw32/x64  
## Running under: Windows 11 x64 (build 26100)  
##   
## Matrix products: default  
## LAPACK version 3.12.1  
##   
## locale:  
## [1] LC\_COLLATE=English\_United Kingdom.utf8   
## [2] LC\_CTYPE=English\_United Kingdom.utf8   
## [3] LC\_MONETARY=English\_United Kingdom.utf8  
## [4] LC\_NUMERIC=C   
## [5] LC\_TIME=English\_United Kingdom.utf8   
##   
## time zone: Europe/London  
## tzcode source: internal  
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] tableone\_0.13.2 psych\_2.5.6 flextable\_0.9.10 ggplot2\_3.5.2   
##   
## loaded via a namespace (and not attached):  
## [1] generics\_0.1.4 fontLiberation\_0.1.0 class\_7.3-23   
## [4] xml2\_1.4.0 lattice\_0.22-7 hms\_1.1.3   
## [7] digest\_0.6.37 magrittr\_2.0.3 evaluate\_1.0.5   
## [10] grid\_4.5.1 RColorBrewer\_1.1-3 fastmap\_1.2.0   
## [13] Matrix\_1.7-3 zip\_2.3.3 e1071\_1.7-16   
## [16] DBI\_1.2.3 survival\_3.8-3 scales\_1.4.0   
## [19] fontBitstreamVera\_0.1.1 textshaping\_1.0.4 labelled\_2.15.0   
## [22] mnormt\_2.1.1 cli\_3.6.5 mitools\_2.4   
## [25] rlang\_1.1.6 fontquiver\_0.2.1 splines\_4.5.1   
## [28] withr\_3.0.2 yaml\_2.3.10 gdtools\_0.4.4   
## [31] tools\_4.5.1 officer\_0.7.0 parallel\_4.5.1   
## [34] uuid\_1.2-1 dplyr\_1.1.4 forcats\_1.0.1   
## [37] vctrs\_0.6.5 R6\_2.6.1 proxy\_0.4-27   
## [40] lifecycle\_1.0.4 ragg\_1.5.0 pkgconfig\_2.0.3   
## [43] pillar\_1.10.2 gtable\_0.3.6 data.table\_1.17.8   
## [46] glue\_1.8.0 Rcpp\_1.1.0 systemfonts\_1.3.1   
## [49] haven\_2.5.5 xfun\_0.53 tibble\_3.3.0   
## [52] tidyselect\_1.2.1 knitr\_1.50 farver\_2.1.2   
## [55] htmltools\_0.5.8.1 nlme\_3.1-168 survey\_4.4-8   
## [58] labeling\_0.4.3 rmarkdown\_2.30 compiler\_4.5.1   
## [61] askpass\_1.2.1 openssl\_2.3.4