

R Example #1

Import the data

Always look at the data after importing!

In most cases, imported data will be a data.frame.

```
cows <- read.csv("C:/hess/STAT511_FA11/RData/CH2_Cows.csv")
str(cows)
```

```
## 'data.frame': 28 obs. of 2 variables:
## $ diet: Factor w/ 2 levels "control","VitA": 1 1 1 1 1 1 1 1 1 1 ...
## $ gain: int 175 132 218 151 200 219 234 149 187 123 ...
```

```
summary(cows)
```

```
##      diet      gain
## control:14  Min.   :123.0
## VitA      :14  1st Qu.:178.2
##           Median :208.5
##           Mean   :211.8
##           3rd Qu.:239.0
##           Max.   :337.0
```

Simple summary stats and graphs (ignoring diet groups)

This is just to get started.

We can use \$ operator to reference a specific column from a data.frame.

```
#Next line gives an error. We need to specify the data!
```

```
#mean(gain)
```

```
mean(cows$gain)
```

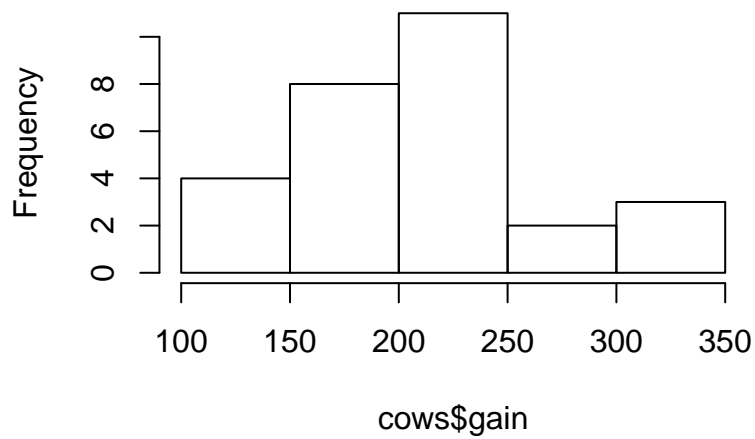
```
## [1] 211.7857
```

```
summary(cows$gain)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    123.0   178.2   208.5   211.8   239.0   337.0
```

```
hist(cows$gain)
```

Histogram of cows\$gain



Summary stats and graphs by diet groups

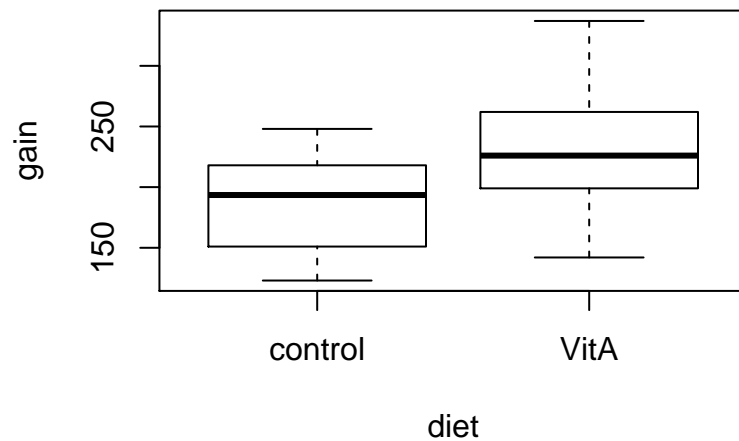
Many functions have a `data =` option that allows us to avoid `$`. When available use the `data =` option!

Note the `~` “squiggle” notation. We will use this many times. Think: $Y \sim X$, Response \sim Predictor.

```
aggregate(gain ~ diet, FUN = mean, data = cows)
```

```
##      diet      gain
## 1 control 187.6429
## 2   VitA 235.9286
```

```
boxplot(gain ~ diet, data = cows)
```



If the `data = option` is not available, can use the `with` function.

```
with(mean(gain), data = cows)
```

```
## [1] 211.7857
```

Tidyverse!

Remember to install packages the first time you use them.

Tidyverse refers to a collection of R packages for data manipulation and plotting. In this example, we use functions from `dplyr` (`summarise`, `group_by`) and `ggplot2` (`qplot`).

The `%>%` “piping” is unique to tidyverse.

```
library(tidyverse)
SumStats <- summarise(group_by(cows, diet),
                      n = n(),
                      mean = mean(gain),
                      sd = sd(gain))
```

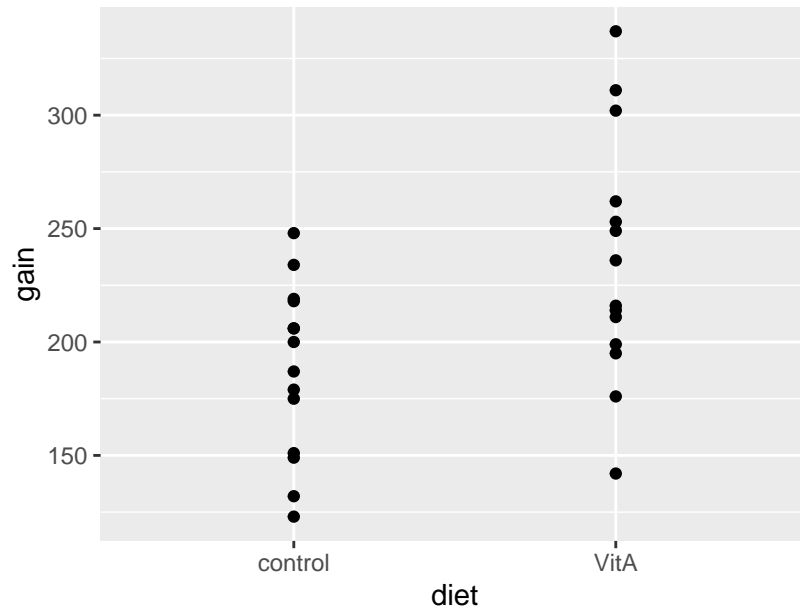
#Equivalently

```
SumStats <- cows %>%
  group_by(diet) %>%
  summarise(n = n(),
            mean = mean(gain),
            sd = sd(gain))
```

SumStats

```
## # A tibble: 2 x 4
##   diet      n mean  sd
##   <fct> <int> <dbl> <dbl>
## 1 control    14  188.  38.1
## 2 VitA       14  236.  54.3
```

```
qplot(x = diet, y = gain, data = cows)
```



```
qplot(x = diet, y = gain, geom = "boxplot", data = cows)
```

