Chapter 3.1 and R introduction

Grethe Hystad

August 16, 2012

How to create and save R scripts

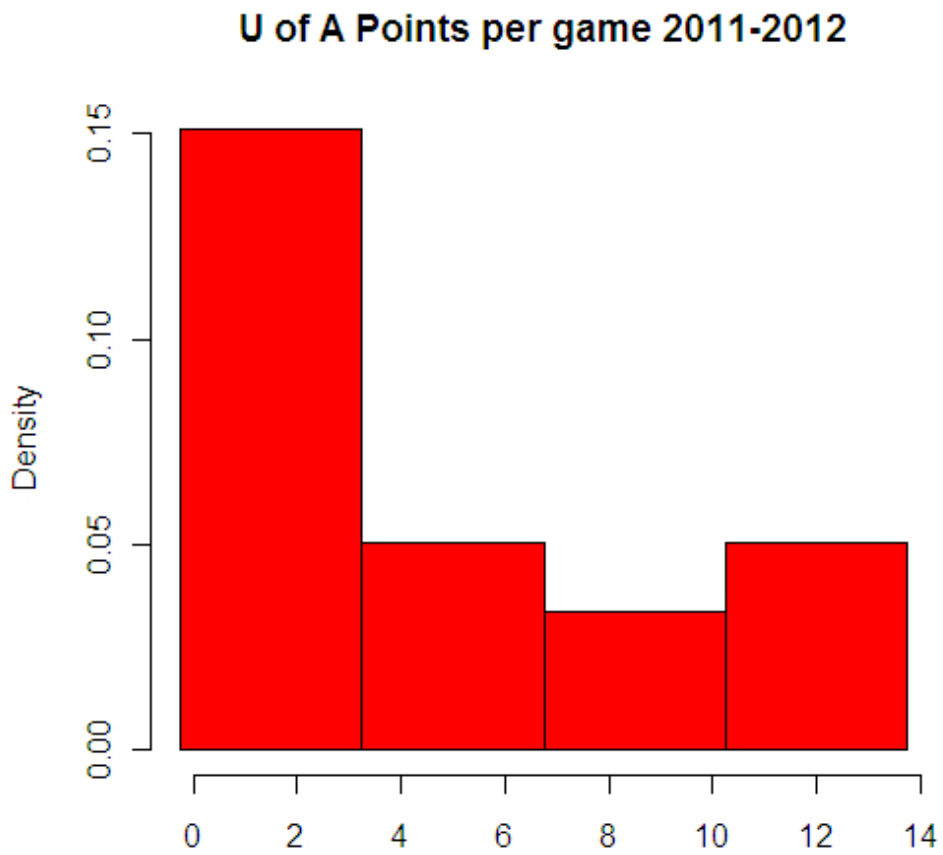
- To create an R script file, click File, then select "New script".
- Type your R code into the R editor that pops up.
- To save it, make the R editor your active window (by clicking on it), then hit File, Save as.
- Later on when you want to open the script, hit File, Open script.
- To run code from the R editor, highlight the lines you want to run, then hit Edit, Run line or selection.

How to read in an excel-file into R

- Save your excel file as a txt file, and file type: Text(MS-DOS), for example test.txt
- To check the location on your file, right-click on the text-file and go to properties.
- In R use the scan function to read in the excel data.
- x=scan(file="C:/Documents and Settings/ghystad/Desktop/Math 361/test.txt")

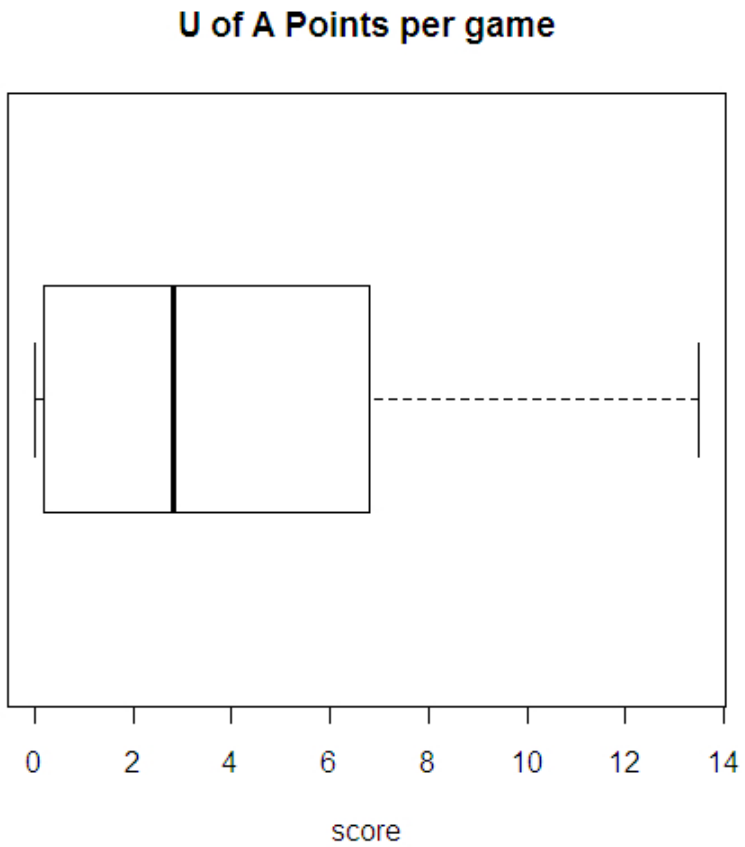
Histogram, UA basketball

The following histogram shows the average points per game for the U of A basketball team in 2011. (Source: AZ Wildcat official site).



boxplot, UA basketball

The following boxplot shows the average score per game for the U of A basketball team in 2011.



UA basketball

C:\Documents and Settings\ghystad\Desktop\Math 361\UofAbaskethist - R Editor

```
scores = scan(file="C:/Documents and Settings/ghystad/Desktop/Math 361/m361UAbasket.txt")
                 # Read in data from Excel-file using scan
scores=c(13.5,13.0,12.2,9.0,6.8,5.4,5.0,4.9,2.8,1.0,1.0,0.3,0.2,0,0,0,0) # create a vector
length(scores)    # length of vector
mean(scores)      # sample mean
sd(scores)        # sample standard deviation
hist(scores)      # frequency histogram
R=max(scores)-min(scores) # range of values
hist(scores,probability=TRUE) # density histogram
rug(jitter(scores))  # add tick marks
hist(scores,breaks=c(-0.25, 3.25, 6.75,10.25,13.75), main="U of A Points per game 2011-2012", xlab="score"
,probability=TRUE, col="red")
                  # with break: specifies the break points
hist(scores,breaks=seq(min(scores)-.25,max(scores)+0.25,3.5)
, main="U of A Points per game 2011-2012", xlab="score",probability=TRUE, col="red")
                  # with break: specifies the range of break points with interval length of 3.5


stem(scores)      # stem-and-leaf charts
stem(scores, scale=3) # Brake up the categories into more groups.
boxplot(scores, main="U of A Points per game",xlab="score", horizontal = TRUE)
summary(scores)
```

UA basketball

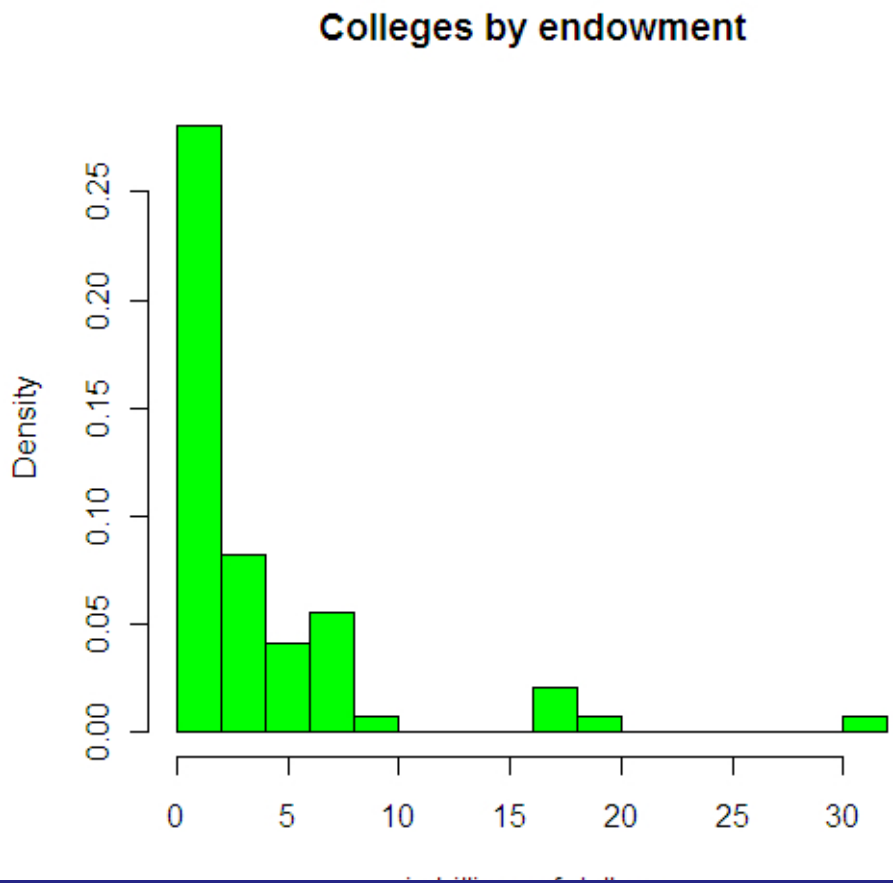
```
> stem(scores)      # stem-and-leaf charts

  The decimal point is 1 digit(s) to the right of the |

  0 | 000000113
  0 | 55579
  1 | 234

> stem(scores, scale=3) # Brake up the categories into more groups.

  The decimal point is at the |

   0 | 000023
   1 | 00
   2 | 8
   3 |
   4 | 9
   5 | 04
   6 | 8
   7 |
   8 |
   9 | 0
  10 |
  11 |
  12 | 2
  13 | 05
```

UA basketball

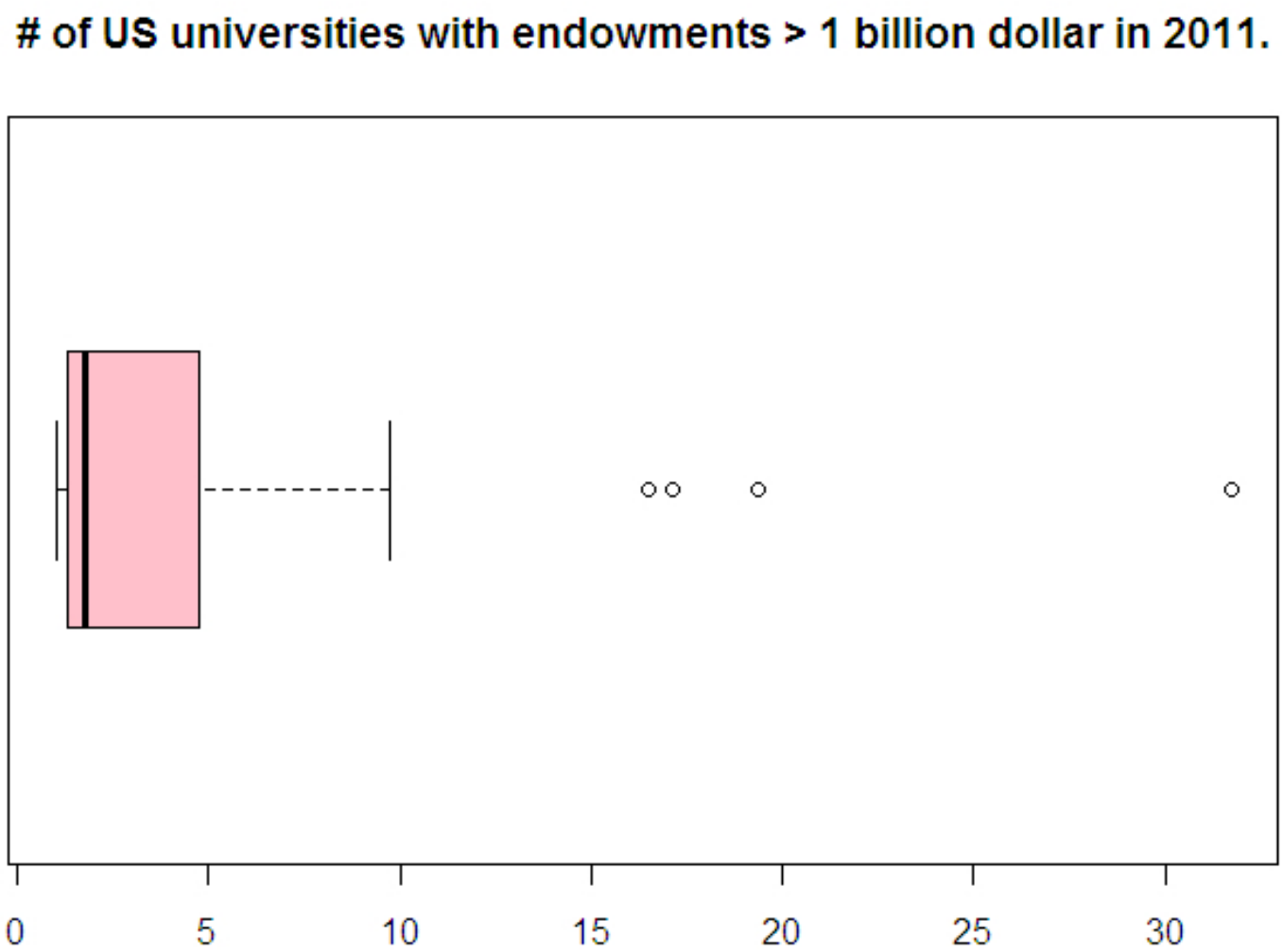
```
> summary(scores)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   0.200   2.800   4.418   6.800  13.500
> 
```

Histogram, Endowment

The following histogram shows the number of US universities with endowments greater than 1 billion dollar in 2011. (Source: Wiki)



Boxplot, Endowment



Endowment

- Skewed to the right.
- One peak
- Mean $= 3.93$ billion
- Median $= 1.75$ billion
- The mean is greater than the median.
- Outliers: more than 16.5 billion dollars.

Endowment

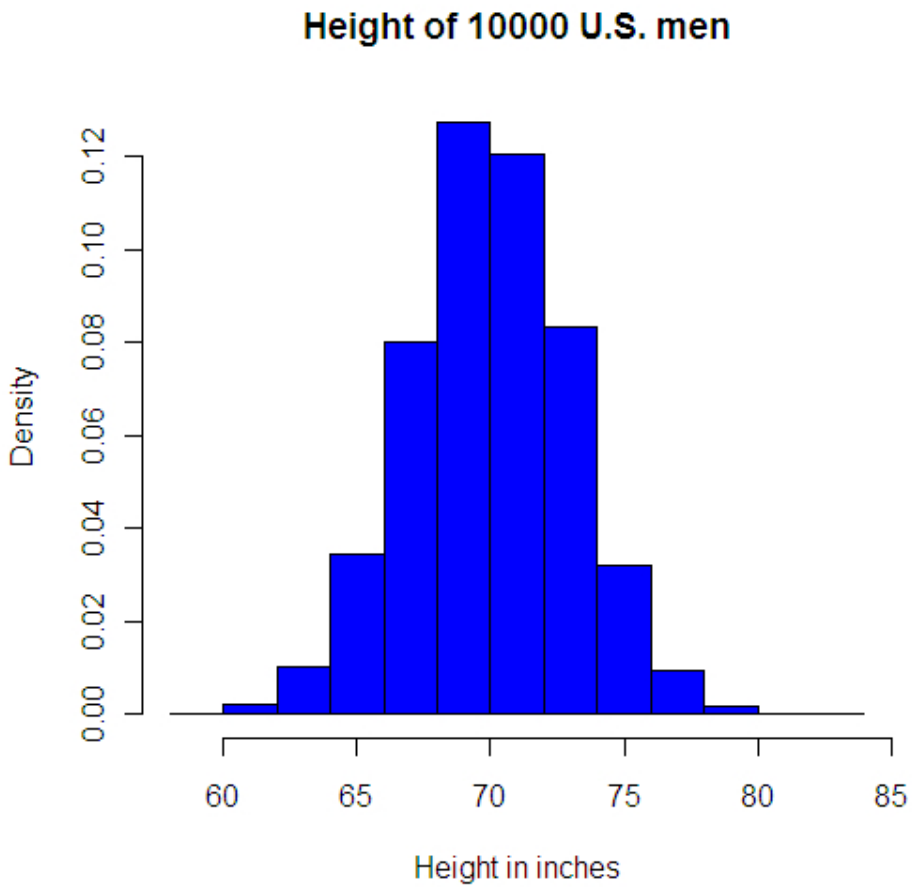
```
x=c(1.004, 1.004, 1.015, 1.017,1.044,1.055,1.058,1.077,1.119,1.125,
1.160,1.160,1.192,1.197,1.212,1.218,1.242,1.250,1.295,1.331,1.400,
1.404,1.430,1.486,1.500,1.500,1.508,1.575,1.601,1.620,1.623,1.642,
1.700,1.703,1.725,1.726,1.746,1.772,1.784,1.873,1.877,2.002,2.121,
2.154,2.261,2.497,2.503,2.527,2.598,2.827,3.413,3.415,3.517,4.451,
4.761,5.059,5.280,5.400,5.747,6.260,6.342,6.575,6.582,7.000,7.183,
7.790, 7.835,9.713,16.503,17.110,17.149,19.374,31.728)
mean(x)
median(x)
hist(x,breaks=12,probability = TRUE,main="Colleges by endowment",
xlab="in billions of dollars", col="green")
summary(x)
q1=quantile(x, 0.25) # 1. quartile
q3=quantile(x, 0.75) # 3. quartile
IQR=q3-q1                  # Interquartile Range
I1=q1-1.5*IQR
I2=q3+1.5*IQR
outliers1=which(x>I2) #locate the outliers
outliers2=which(x<I1)
length(outliers1)          # count the number of outliers
length(outliers2)
boxplot(x,horizontal=TRUE,main="# of US universities with endowments > 1 billion dollar in 2011."
,col="pink")
```

Endowment

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.004   1.295   1.746   3.927   4.761  31.730
> q1=quantile(x, 0.25) # 1. quartile
> q3=quantile(x, 0.75) # 3. quartile
> IQR=q3-q1                  # Interquartile Range
> I1=q1-1.5*IQR
> I2=q3+1.5*IQR
> outliers1=which(x>I2) #locate the outliers
> outliers1
[1] 69 70 71 72 73
> outliers2=which(x<I1)
> outliers2
integer(0)
> length(outliers1)          # count the number of outliers
[1] 5
> length(outliers2)
[1] 0
```

Histogram, height

The following histogram shows the height of 10000 men in centimeter.



Height

- Mean = 70 inches.
- Median = 70 inches.
- Symmetric
- One peak