|  |  |  |
| --- | --- | --- |
| MATH 1350 | **Statistics for Information Technology** |  |

**Lab # 1 – Exploratory Data Analysis With R**

Answer/Grading Sheet

| **Step:** | **Answer (if requested)** | **Mark** |  |
| --- | --- | --- | --- |
| 3 | * load the mosaic package |  | /1 |
| 4 | * import data files * change working directory |  | /2  /1 |
| 5 | * copy/save document with the name *nnn*Lab1Answers.docx |  | /1 |
| 6 | Answer to Question 1:  From the data set, I see that there is a large difference in download times between both data sets because the server 1 seems to have much higher and lower extremes, whereas server 2 values seems to have less deviation. |  | /3 |
| 7 | * server 1 & server 2 histograms (Export -> Save as Image in R and then Insert -> Pictures in Word)   Answer to Question 2:  Server1Server 2  From the histograms we can tell that there is a large difference in the density between server 1 and 2.  Answer to Question 3:  Based on these histograms I would say that the majority users would not notice a large different between the two servers as on both graphs the time tends to peak around 5-10 seconds |  | /2  /1  /2 |
| 8 | * server 1 & server 2 log(Time) histograms (paste them in here)   Answer to Question 4:  Server 1  Server 2  The shape is a bell curve  Answer to Question 5:  The main differences between these two graphs are the differences in density and extremes. Server 1 has a lower density and higher extremes whereas server 2 has higher density with lower extremes |  | /2  /2  /2 |
| 9 | * mean time for server 1 & server 2: (type them in here)   Server 1: 12.69262  Server 2: 7.583346  Answer to Question 6:  They seem to be measuring the average time it takes to download a 50mb file.  Answer to Question 7:  Yes, with server 2 having a higher density it would make sense for it to have a lower average time. In other words, the are more download times that are between 5-10 seconds and lesser extremes compared to server 1 so it would make sense for server 2 for have a lower average time. |  | /2  /2  /2 |
| R script  Paste your R script here. It should contain **all** the commands you used to complete the lab. Commands entered in the console should be copied into your script!  server1data <- read.delim("C:/Users/Markus/OneDrive - BCIT/Desktop/Term2/MATH 1350 Statistics for IT/Week1/server1data.txt", comment.char="#")  ?read.delim  setwd("C:/Users/Markus/OneDrive - BCIT/Desktop/Term2/MATH 1350 Statistics for IT/Week1")  Server2Data=read.delim("server2data.txt",header=TRUE,sep="\t",na.strings="NA",comment.char="#")  histogram(~log(Time), data=server1data)  histogram(~log(Time), data=Server2Data)  mean(~Time, data=server1data)  mean(~Time, data=Server2Data) | |  | /5 |

/30