

NAME: Markus Afonso

Set: C \_\_\_\_\_

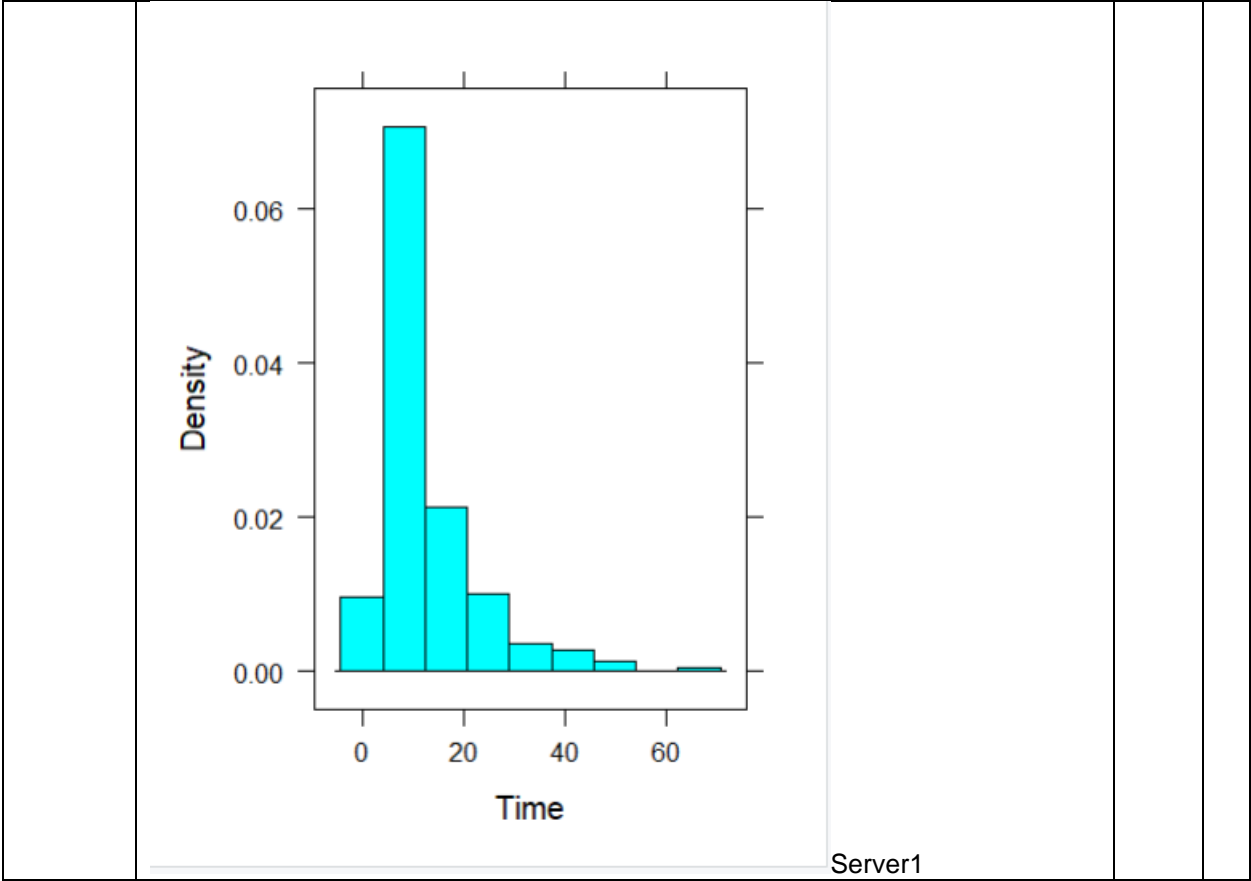
MATH 1350

Statistics for Information Technology

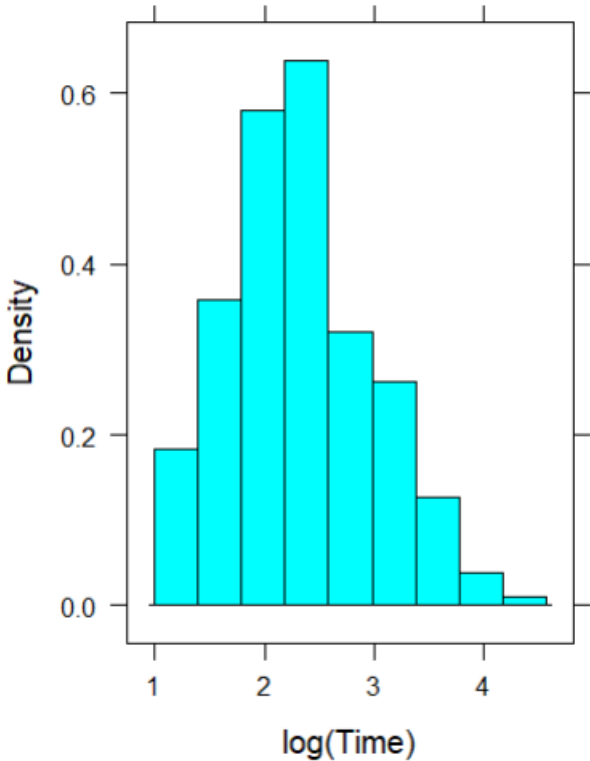
## Lab # 1 – Exploratory Data Analysis With R

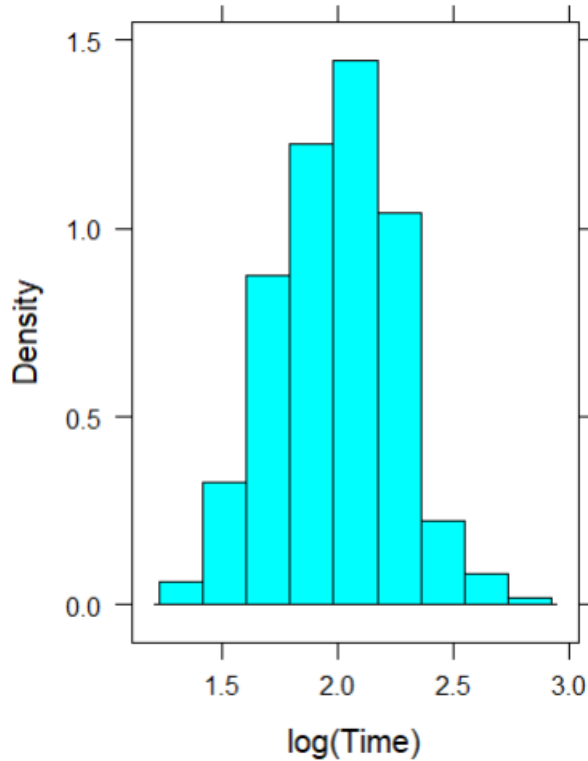
## Answer/Grading Sheet

Step:	Answer (if requested)	Mark	
3	<ul style="list-style-type: none"> <li>load the mosaic package</li> </ul>		/1
4	<ul style="list-style-type: none"> <li>import data files</li> <li>change working directory</li> </ul>		/2 /1
5	<ul style="list-style-type: none"> <li>copy/save document with the name <i>nnnLab1Answers.docx</i></li> </ul>		/1
6	<p>Answer to Question 1:</p> <p>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.</p>		/3
7	<ul style="list-style-type: none"> <li>server 1 &amp; server 2 histograms (Export -&gt; Save as Image in R and then Insert -&gt; Pictures in Word)</li> </ul> <p>Answer to Question 2:</p>		/2 /1 /2



Step:	Answer (if requested)	Mark																							
	<div><table><caption>Data for Server 2 Histogram</caption><thead><tr><th>Time Interval</th><th>Density</th></tr></thead><tbody><tr><td>[2, 3]</td><td>0.01</td></tr><tr><td>[3, 4]</td><td>0.00</td></tr><tr><td>[4, 5]</td><td>0.10</td></tr><tr><td>[5, 6]</td><td>0.19</td></tr><tr><td>[6, 7]</td><td>0.17</td></tr><tr><td>[7, 8]</td><td>0.10</td></tr><tr><td>[8, 9]</td><td>0.02</td></tr><tr><td>[9, 10]</td><td>0.01</td></tr><tr><td>[10, 11]</td><td>0.005</td></tr><tr><td>[11, 12]</td><td>0.002</td></tr></tbody></table></div> <p>Server 2</p> <p>From the histograms we can tell that there is a large difference in the density between server 1 and 2.</p> <p>Answer to Question 3:</p> <p>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</p>	Time Interval	Density	[2, 3]	0.01	[3, 4]	0.00	[4, 5]	0.10	[5, 6]	0.19	[6, 7]	0.17	[7, 8]	0.10	[8, 9]	0.02	[9, 10]	0.01	[10, 11]	0.005	[11, 12]	0.002		
Time Interval	Density																								
[2, 3]	0.01																								
[3, 4]	0.00																								
[4, 5]	0.10																								
[5, 6]	0.19																								
[6, 7]	0.17																								
[7, 8]	0.10																								
[8, 9]	0.02																								
[9, 10]	0.01																								
[10, 11]	0.005																								
[11, 12]	0.002																								
8	<ul style="list-style-type: none"><li>server 1 &amp; server 2 log(Time) histograms (paste them in here)</li></ul> <p>Answer to Question 4:</p>		<div>/</div> <div>2</div> <div>/</div> <div>2</div> <div>/</div> <div>2</div>																						

Step:	Answer (if requested)	Mark																					
	<div><p>A histogram showing the density of log(Time) for Server 1. The x-axis is labeled 'log(Time)' and ranges from 1 to 4. The y-axis is labeled 'Density' and ranges from 0.0 to 0.6. The histogram consists of 10 bars of width 0.5. The distribution is unimodal and slightly right-skewed, with a peak density of approximately 0.64 at log(Time) = 2.5.</p><table><tr><th>log(Time) Bin</th><th>Density</th></tr><tr><td>1.0 - 1.5</td><td>0.18</td></tr><tr><td>1.5 - 2.0</td><td>0.36</td></tr><tr><td>2.0 - 2.5</td><td>0.58</td></tr><tr><td>2.5 - 3.0</td><td>0.64</td></tr><tr><td>3.0 - 3.5</td><td>0.32</td></tr><tr><td>3.5 - 4.0</td><td>0.26</td></tr><tr><td>4.0 - 4.5</td><td>0.12</td></tr><tr><td>4.5 - 5.0</td><td>0.04</td></tr><tr><td>5.0 - 5.5</td><td>0.01</td></tr></table></div> <p>Server 1</p>	log(Time) Bin	Density	1.0 - 1.5	0.18	1.5 - 2.0	0.36	2.0 - 2.5	0.58	2.5 - 3.0	0.64	3.0 - 3.5	0.32	3.5 - 4.0	0.26	4.0 - 4.5	0.12	4.5 - 5.0	0.04	5.0 - 5.5	0.01		
log(Time) Bin	Density																						
1.0 - 1.5	0.18																						
1.5 - 2.0	0.36																						
2.0 - 2.5	0.58																						
2.5 - 3.0	0.64																						
3.0 - 3.5	0.32																						
3.5 - 4.0	0.26																						
4.0 - 4.5	0.12																						
4.5 - 5.0	0.04																						
5.0 - 5.5	0.01																						

Step:	Answer (if requested)	Mark	
	 <p style="text-align: right;">Server 2</p> <p>The shape is a bell curve</p> <p>Answer to Question 5:</p> <p>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</p>		
9	<ul style="list-style-type: none"> <li>mean time for server 1 &amp; server 2: (type them in here)</li> </ul> <p>Server 1: 12.69262 Server 2: 7.583346</p> <p>Answer to Question 6:</p> <p>They seem to be measuring the average time it takes to download a 50mb file.</p> <p>Answer to Question 7:</p> <p>Yes, with server 2 having a higher density it would make sense for it to have a lower average time. In other words, there are more download times that are</p>		/ 2 / 2 / 2

Step:	Answer (if requested)	Mark	
	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.		
R script	<p>Paste your R script here. It should contain <b>all</b> the commands you used to complete the lab. Commands entered in the console should be copied into your script!</p> <pre> server1data &lt;- 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) </pre>		/5