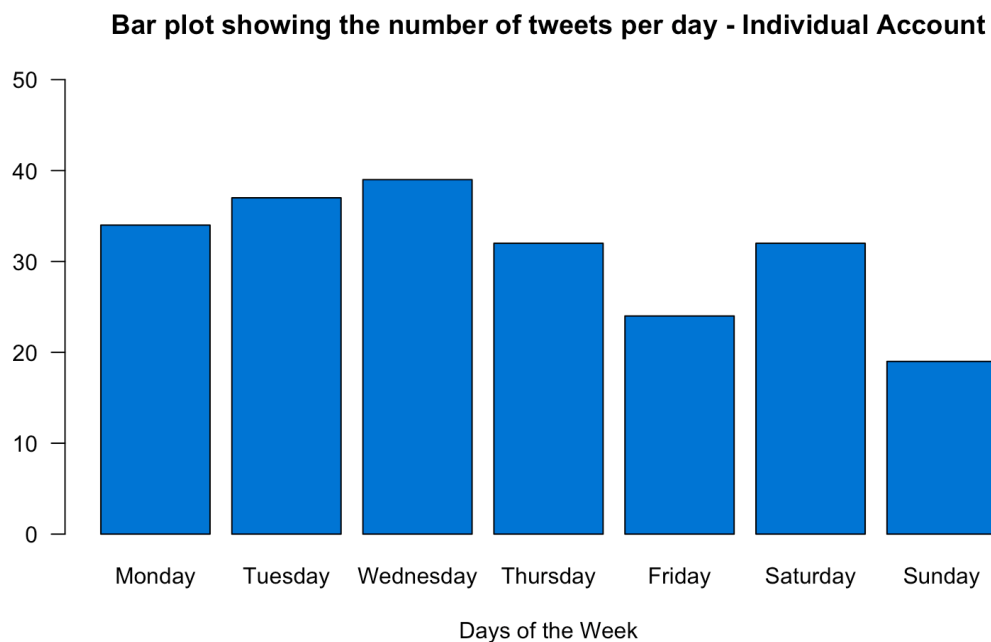


## Analysis 1

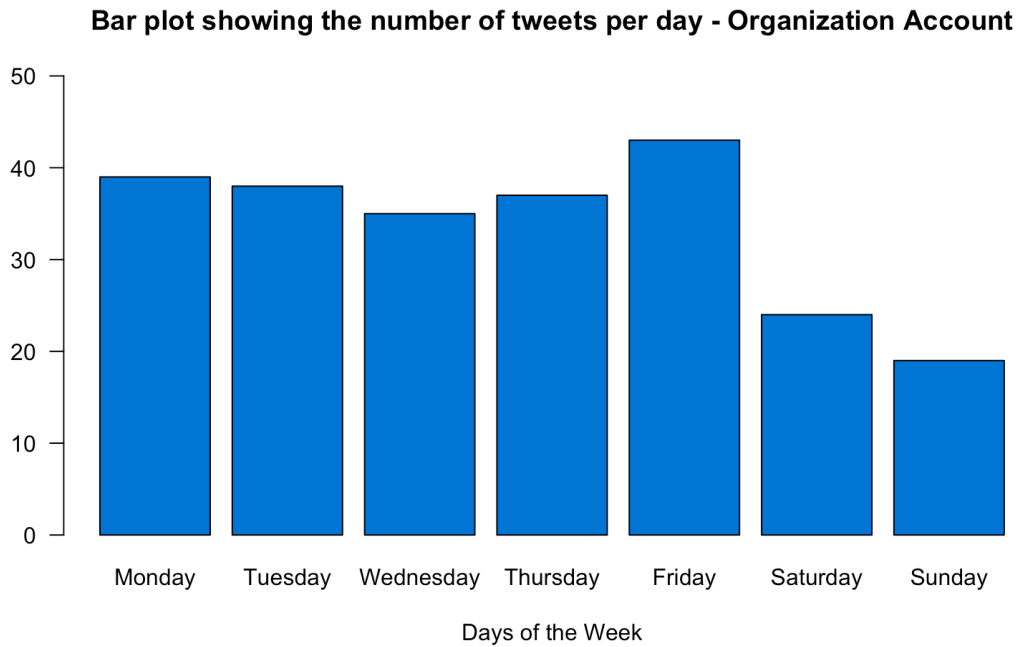
I will be analyzing @theJagmeetSingh from my personal accounts and @UWaterloo from my organizational accounts.

Day	Personal account	Organizational sccount
Monday	34	39
Tuesday	37	38
Wednesday	39	35
Thursday	32	37
Friday	24	43
Saturday	32	24
Sunday	19	19

Barplots of day . of . week:



Barplot for the weekly twitter summary of @theJagmeetSingh



Barplot for the weekly twitter summary of @UWaterloo

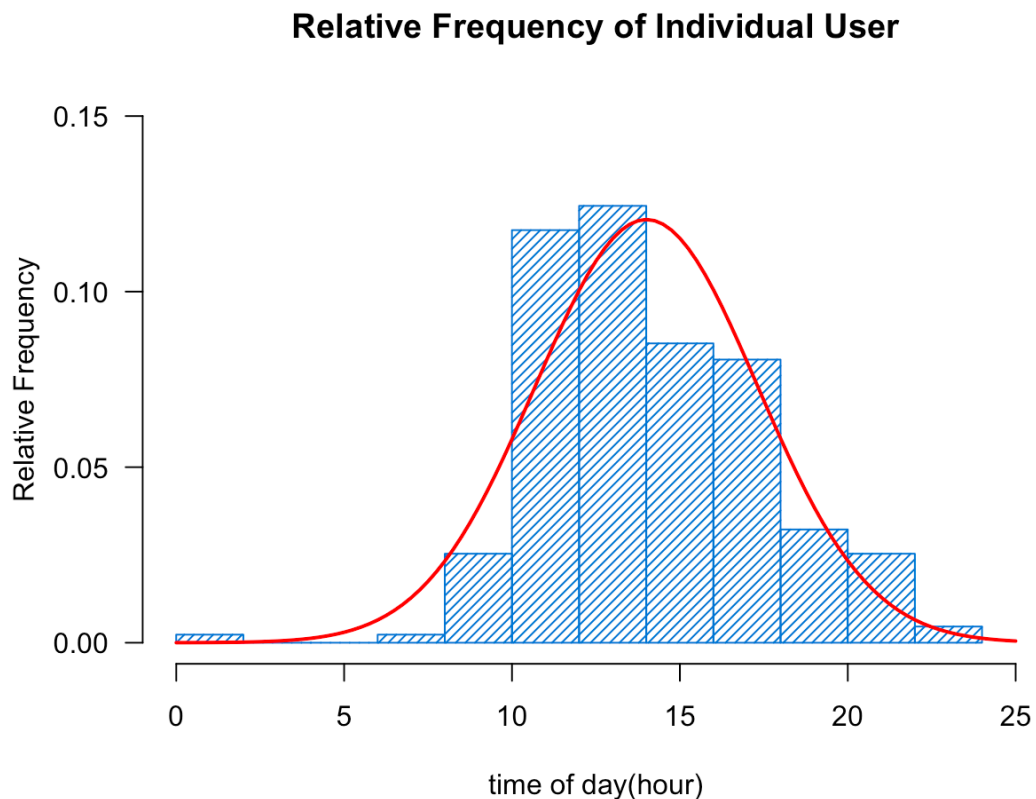
The distributions of `day.of.week` for @theJagmeetSingh and @UWaterloo are somewhat similar. For @theJagmeetSingh, we can see that the account posts regularly throughout the week with the exception of Sunday, while for @UWaterloo, we can see that the account posts regularly throughout the week with the exception of Saturday and Sunday. Both accounts belong to very public entities, we can see that they post very often throughout the week. The decline in tweets by the organizational account suggests the presence of a chief communications officer or someone similar, a person who is in charge of handling the organization's Twitter account.



I will be analyzing @theJagmeetSingh from my personal accounts and @UWaterloo from my organizational accounts.

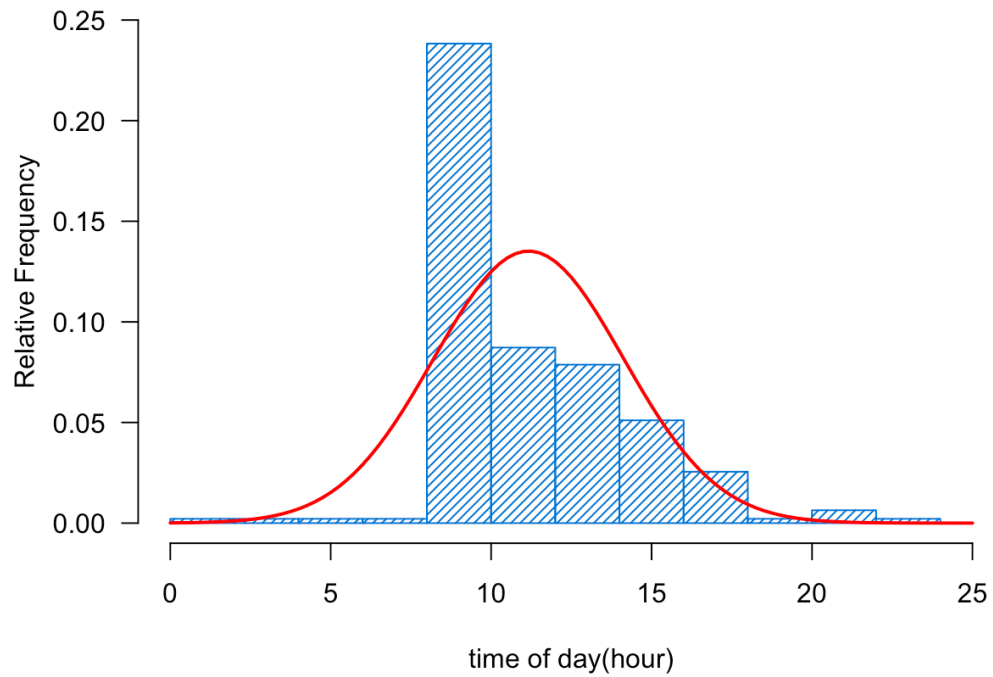
Sample statistic	Personal account	Organizational account
Mean	14.0040	11.169
Median	13.3339	10.001
SD	3.309865	2.951036
Skewness	0.2106931	0.9231687
Kurtosis	3.530165	4.480478

Relative frequency histograms of time.of.day.hour with superimposed probability density function curves:

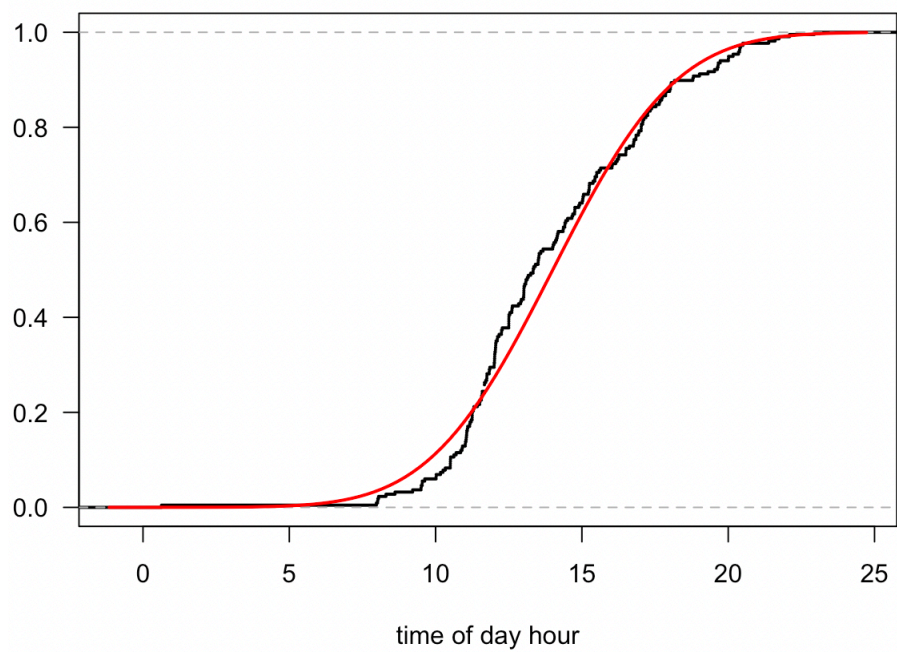


Relative Frequency histogram of @theJagmeetSingh

**Relative Frequency of Organization User**

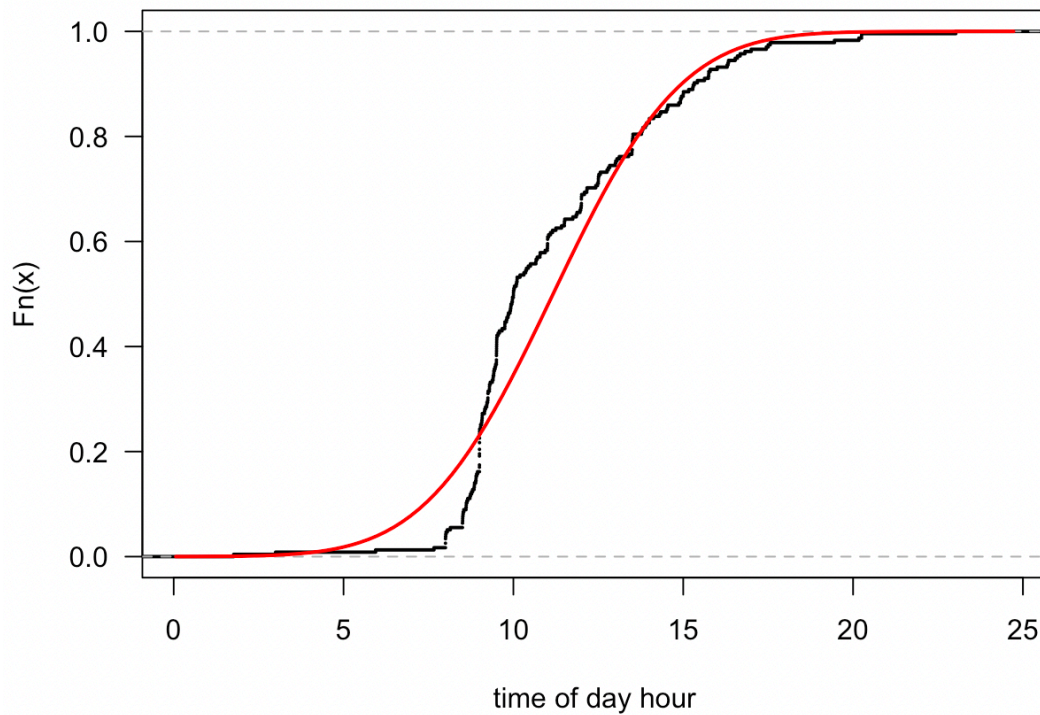


**e.c.d.f. of time of day hour for Individual User**



ECDF of time of day (in hours) for @theJagmeetSingh

### e.c.d.f. of time of day hour for Organization User



ECDF of time of day (in hours) for @UWaterloo

Empirical cumulative distribution function plots of `time.of.day.hour` with superimposed cumulative distribution function curves:

In my sample, @theJagmeetSingh has tweeted approximately 10% of their tweets before 10:30 am, and @UWaterloo has tweeted approximately 10% of their tweets before 8:37 am.

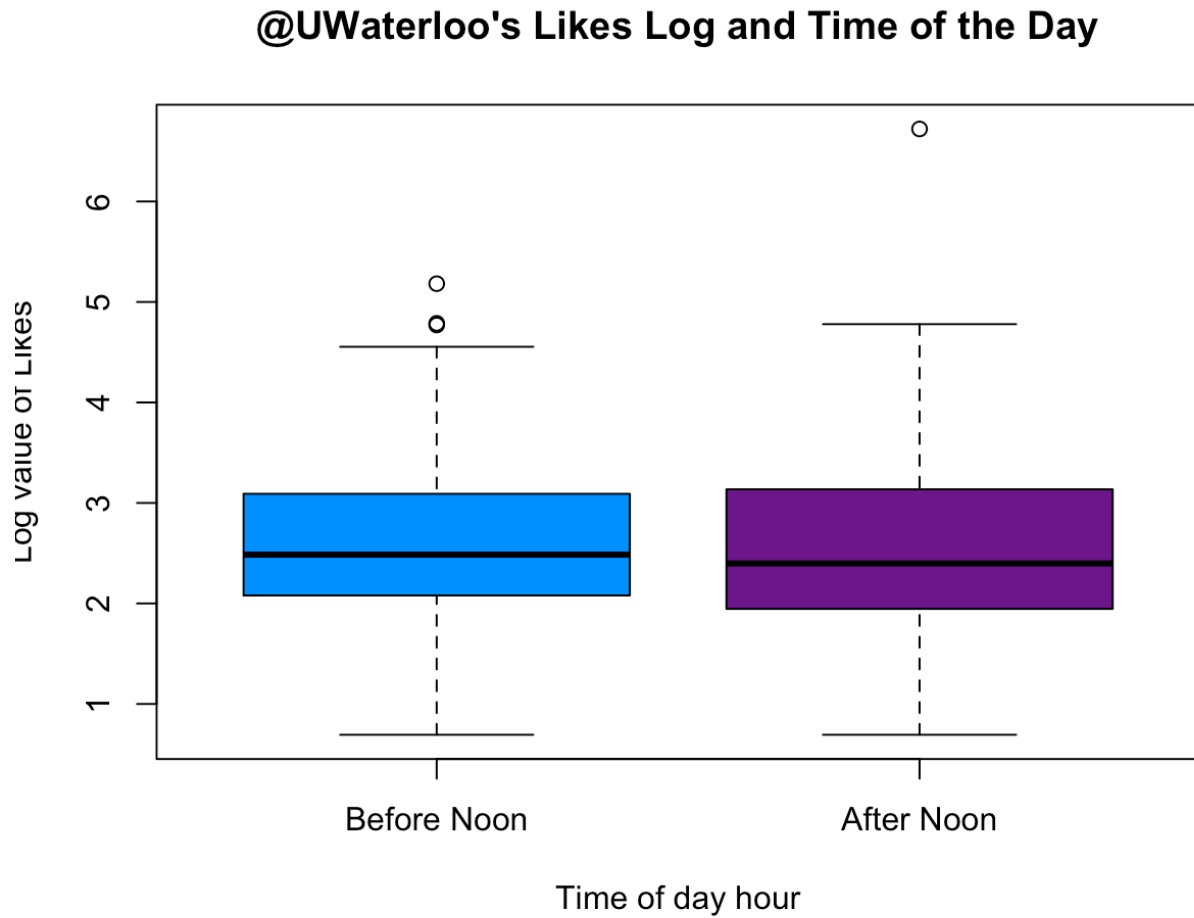
**Personal account (@theJagmeetSingh):** Based on the plot in part 2c, we can see that the distribution is a slight right-skew and there is a slight presence of outliers in the graph. This is further proved by the slightly positive skewness and a kurtosis value of greater than 3. While for data generated from a Gaussian distribution we would expect to see a skewness value of 0 and a kurtosis value of 3. We can further see that the mean and median are very close but not equal as is the case with a Gaussian distribution. Overall, the Gaussian model fits very well with the distribution as the value of kurtosis is close to 3, the value of skewness is close to 0 and the mean and median are very close to one another.

**Organizational account (@UWaterloo):** Based on the plot in part 2c, we can see that the distribution has a right-skew and there is the presence of outliers in the graph. This is symbolized by the positive value of skewness and a value of kurtosis that is greater than 3. While for data generated from a Gaussian distribution we would expect to see a skewness equal to 0 and a kurtosis equal to 3. We further see that the mean and the median are not close in contradiction to

the case of the Gaussian distribution. Overall, the Gaussian model does not fit properly with the distribution.

I will be analyzing @UWaterloo for Analysis 3.

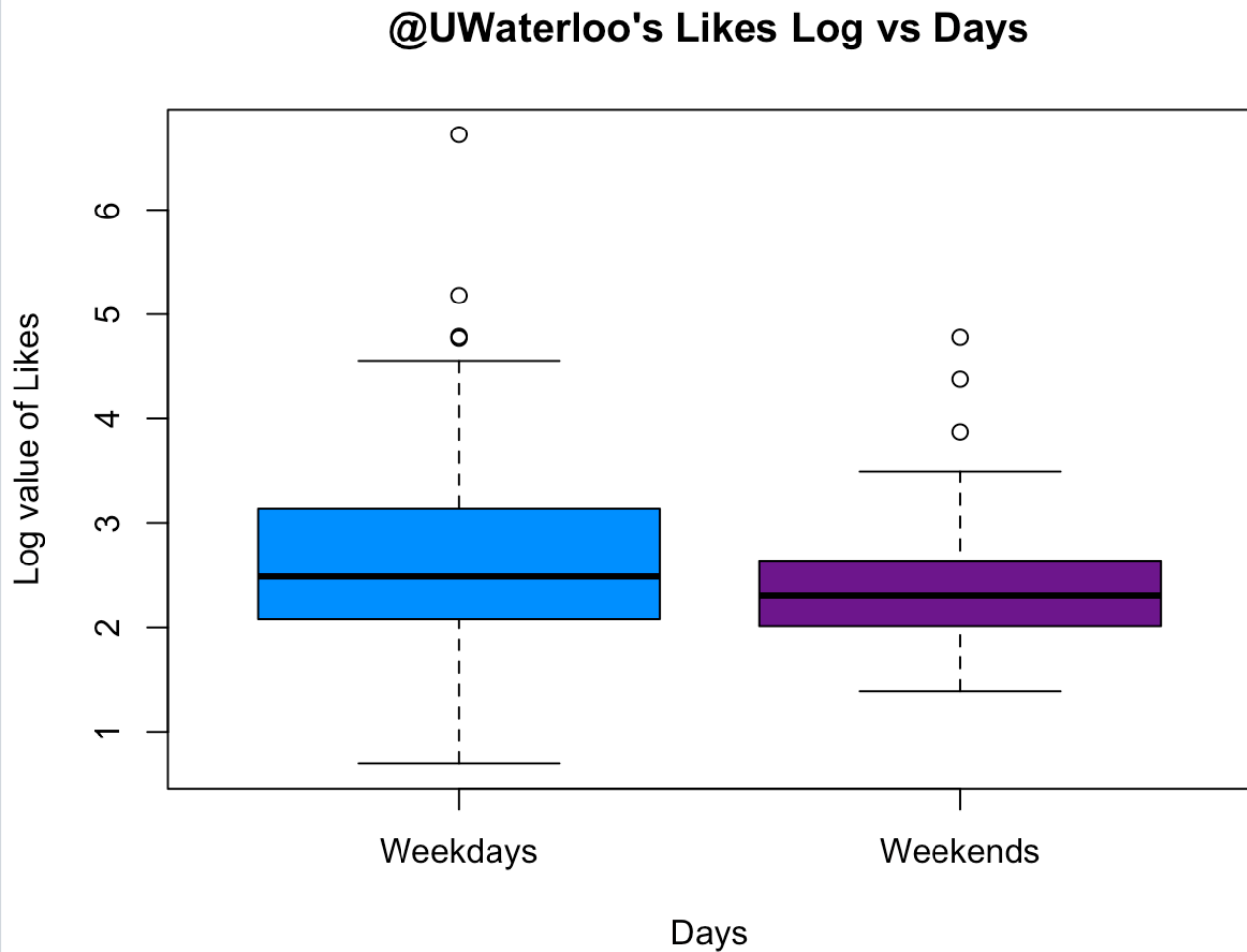
Side-by-side boxplot of `likes` for tweets published before noon vs. afternoon:



@UWaterloo's boxplot that compares `likes.log` for tweets that are published before noon to tweets published at or after noon



Side-by-side boxplot of likes for tweets published on weekends vs. weekdays:



@UWaterloo's boxplot that compares likes.log for tweets that are published before noon to tweets published on Weekdays and Weekends

Based on the results of Analysis 3b, we would recommend publishing tweets on or after noon that are liked the most, this is because the boxplot for before noon has a smaller IQR and a smaller upper range than compared to the boxplot for the afternoon [This is because most of the people connected to the organization are extremely occupied in the morning due to classes or other early morning activities, this reduces their ability to check their Twitter]. Based on the results of Analysis 3c, we would recommend publishing tweets during the weekdays rather than

at the weekends, because according to the boxplot the IQR and the range of likes are much larger on the weekdays rather than at the weekends.