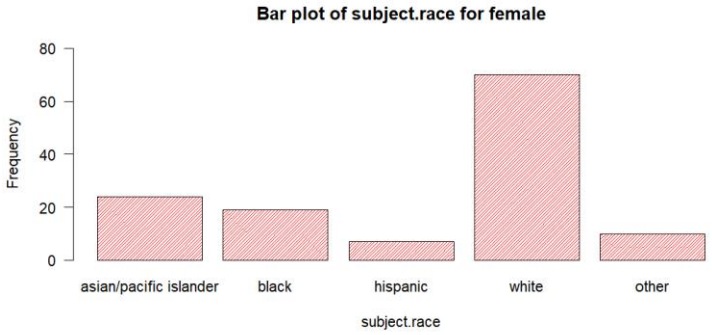
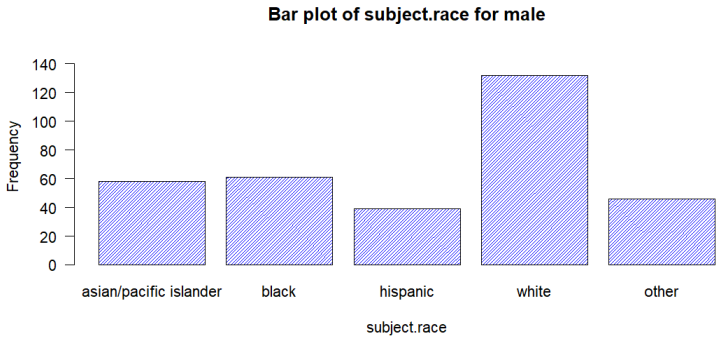
The following below is my data analysis the subject.race variate for San Francisco.

Race	Female Frequency (%)	Male Frequency (%)
Asian/Pacific Islander	24(18.5%)	58(17.3%)
Black	19(14.6%)	61(18.2%)
Hispanic	7(5.4%)	39(11.6%)
White	70(53.9%)	132(39.3%)
Other	10(7.7%)	46(13.7%)

The graph below is Bar plots of my chosen variate:



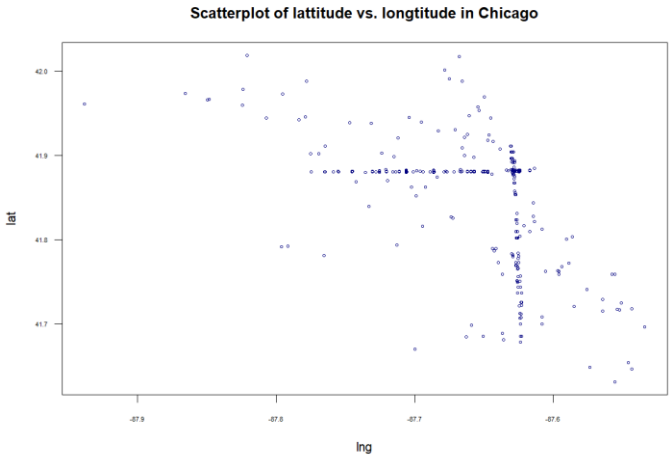


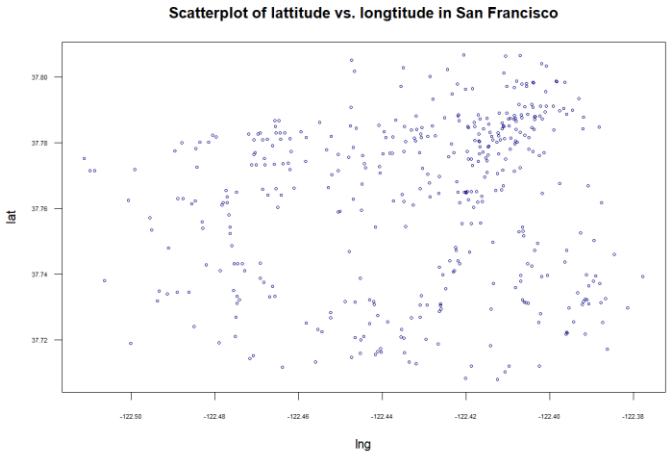
The distributions of subject.race for subjects identified as female and male are somewhat similar. For subjects identified as female, we can see that Whites have a frequency close to 80, far more than the other four categories(all less than 30), while for subjects identified as male, we can see that Whites have a frequency close to 140, also more than the other four categories(all less than 60), The only difference is that in the subject.race for male graph, the other four categories (asian/pacific islander", "black", "hispanic", "other") make up a larger percentage of the population overall, and the frequency gap with Whites is not as pronounced.

		subject.sex	
		Female	Male
City	Chicago	137	280
	San Francisco	130	336

The proportion of traffic stops in Chicago which were of subjects identified as female was 0.329. The proportion of traffic stops in San Francisco which were of subjects identified as female was 0.279.

The relative risk is calculated by (The proportion of traffic stops in Chicago which were of subjects identified as female)/(The proportion of traffic stops in San Francisco which were of subjects identified as female). This gives a relative risk of 1.178.



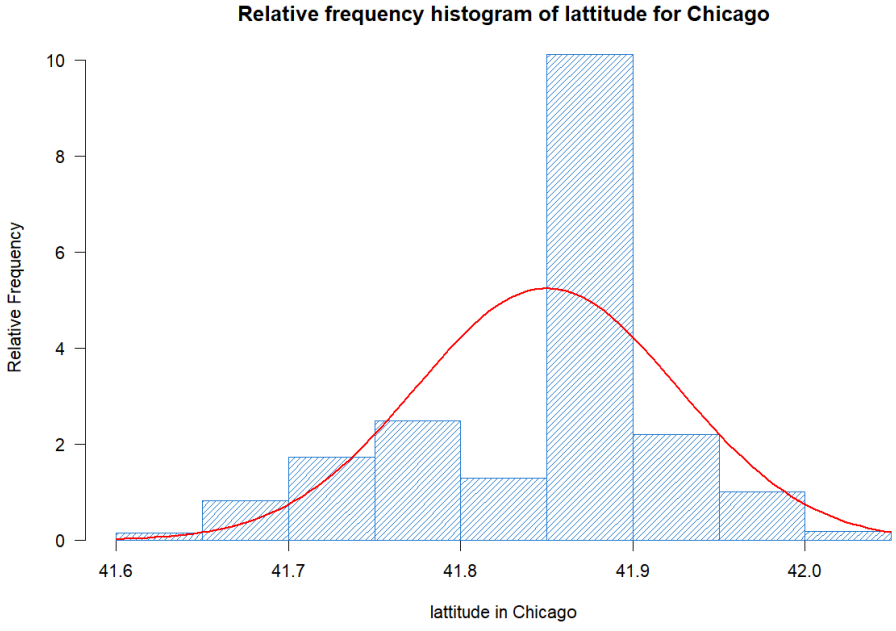


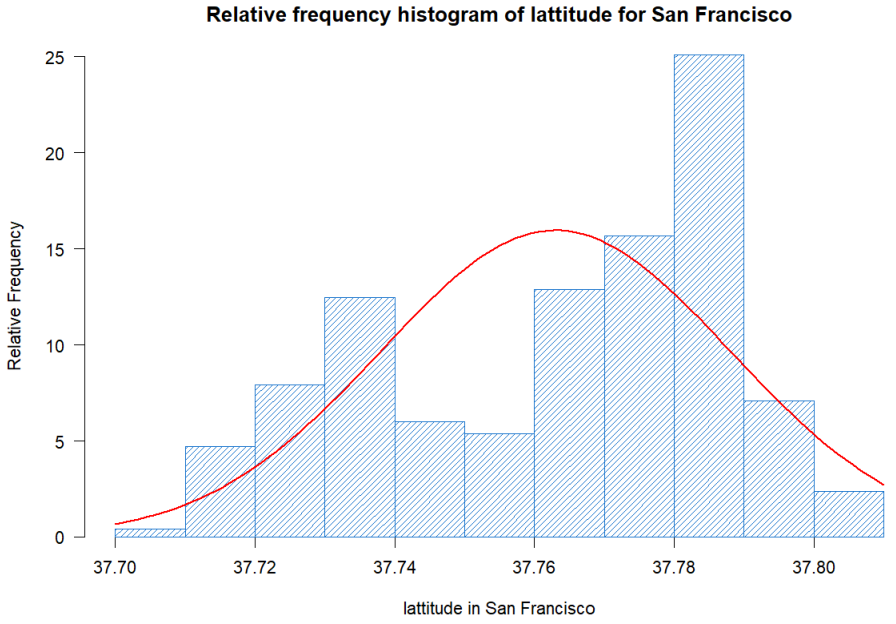
The sample correlation between latitude and longitude for Chicago is -0.5812. This suggests that as latitude increases, longitude tends to decrease, and vice versa (negative sign). And the magnitude of the correlation coefficient (-0.5812) suggests a **moderate linear relationship** of this negative relationship.

The sample correlation between latitude and longitude for San Francisco is 0.1320. This suggests that as latitude increases, longitude tends to increase as well, and vice versa (negative sign). And the magnitude of the correlation coefficient (0.1320) suggests a **weak linear relationship** of the positive relationship.

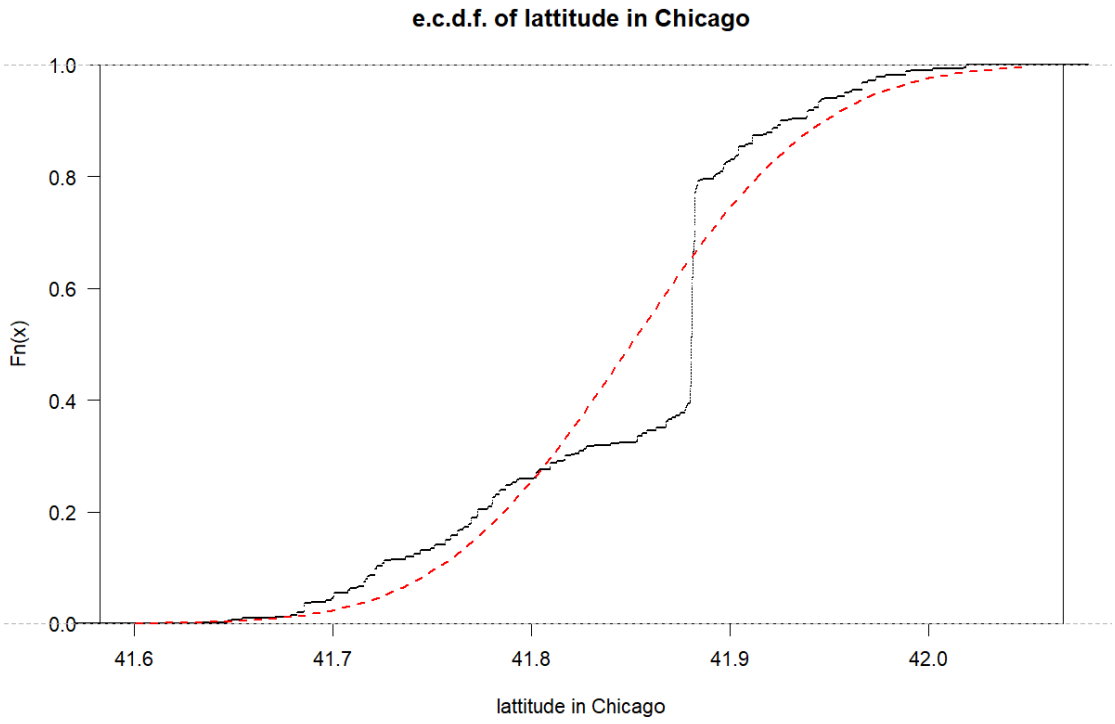
Sample statistic	Chicago	San Francisco
Mean	41.850	37.763
Median	41.880	37.770
SD	0.076	0.025
Skewness	-0.685	-0.461
Kurtosis	2.883	2.019

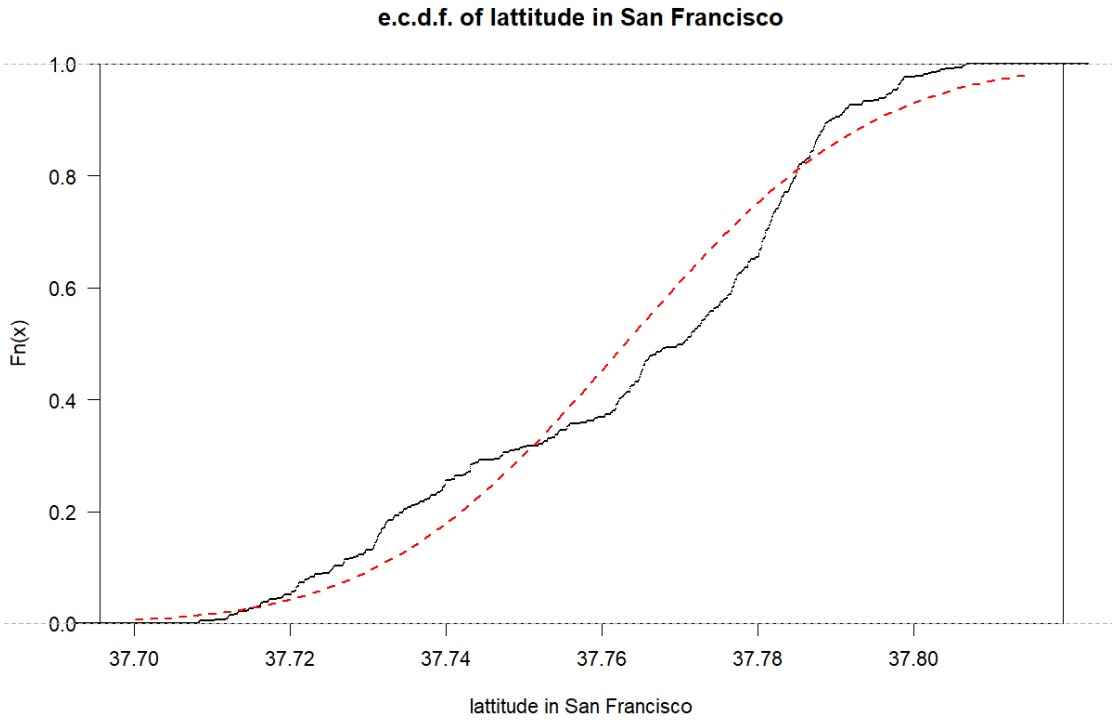
Relative frequency histograms of lat for each city with superimposed probability density function curves:





The graph below is Empirical cumulative distribution function plots of lat for each city with superimposed cumulative distribution function curves:





Chicago:

Based on the plot in Analysis 2e, we can see that the relative frequency peaks in the interval of latitude from 41.85 to 41.9 (10), which is much higher than the other columns, and shows a sudden decrease in the interval from 41.8 to 41.85. While for data generated from a Gaussian distribution we would expect to see the relative frequency is axisymmetric, the axis of symmetry is latitude=41.85, and the corresponding frequency is about 5.

In the e.c.d.f. plot of 2f, we can see that at latitude about 41.85, there is a steep increase in the image of Fn(x), which goes from 0.4 to 0.8 in a straight line. At the same time, it can also be seen from the image that the black line does not match the red line(Gaussian model for ecdf) severely in the region from 41.8 to 41.9.

The skewness of latitude for Chicago is -0.685, and the Kurtosis for latitude for Chicago is 2.883. A negative skewness and positive kurtosis suggest that the latitude values are clustered more to the right of the mean, possibly indicating that Chicago tends to have relatively consistent latitude values with some outliers or extreme values.

Overall, the Gaussian model doesn't fit well with the data observed in Chicago.

San Franciso:

Based on the plot in Analysis 2e, we can see that the latitude shows a gradual increase in the range of 37.70~37.74 and 37.75~37.79, however, it shows a sudden and drastic decrease around the points of 37.74 and 37.79. While for data generated from a Gaussian distribution we would expect to see The image shows symmetry at the place latitude=37.76 and the relative frequency at the highest point is about 15.

In the e.c.d.f. plot of 2f, we can see that the red line has a different pattern than the black line, which is a hair above the red line at latitude less than 37.75. the black line is lower than the red line when latitude lies between 37.75 and 37.78.

Overall, the Gaussian model doesn't fit well with the data observed in San Franciso.