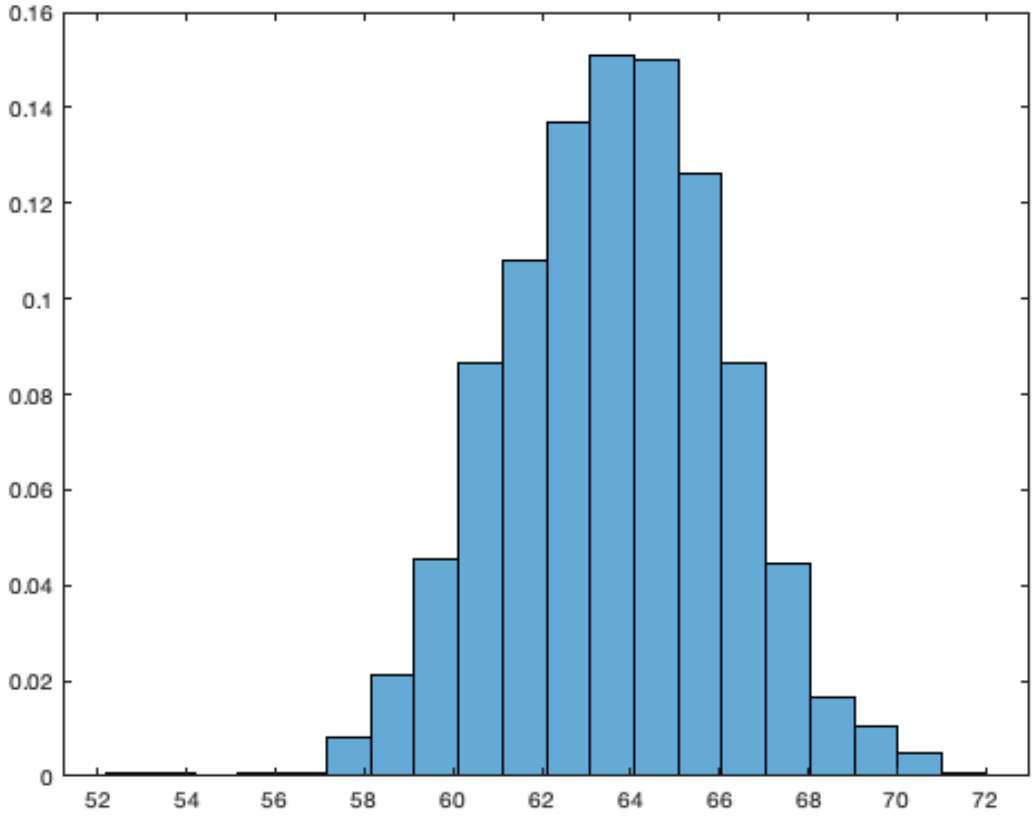
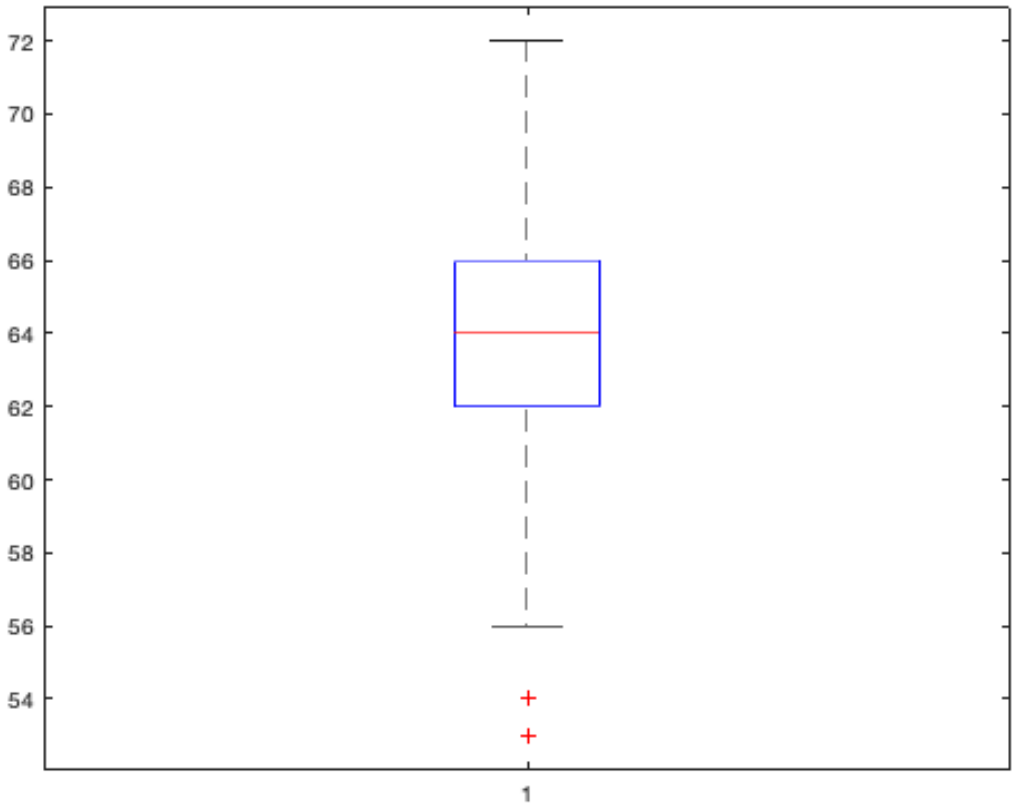
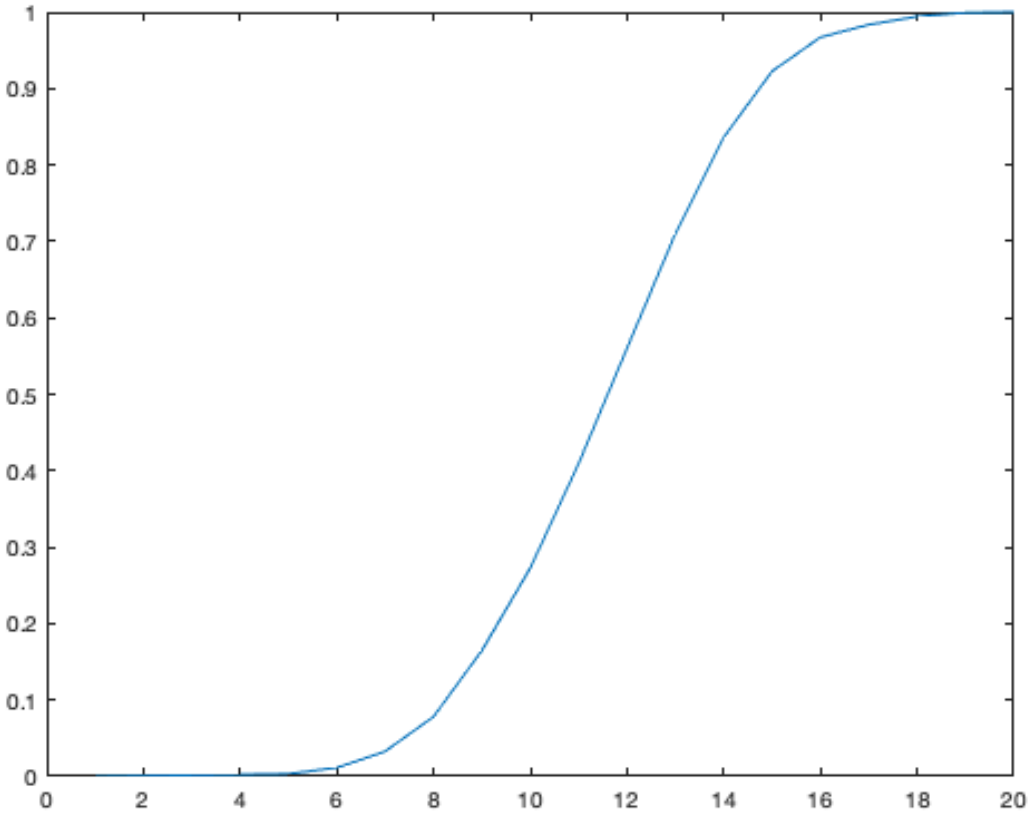
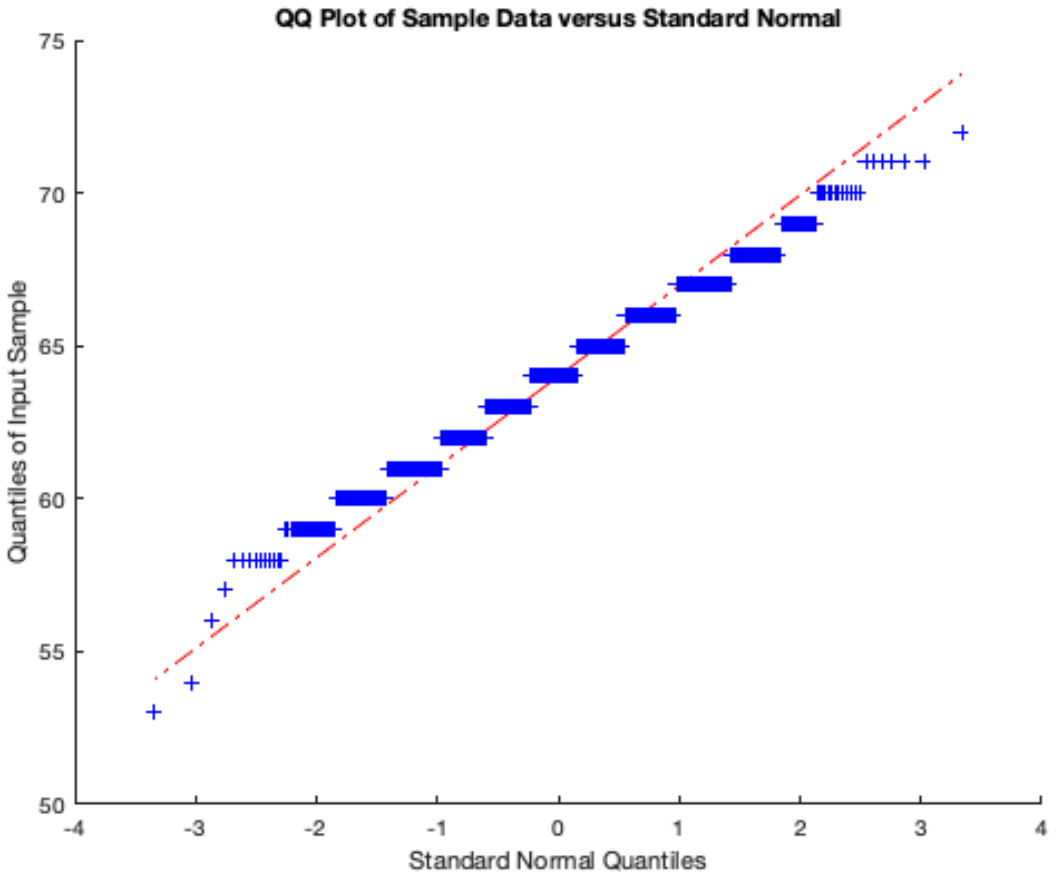
```
% Ben Juarez
% PS1Q5 - Playing With Data Summaries

% Part a
birth = importdata("birth.txt");
mother_heights = birth(:,5);
mother_heights = mother_heights(mother_heights<99); % removing 99 unknown
histogram(mother_heights, 20, "Normalization", "probability")
snapnow
% Around 20 bins appears to be optimal for representing the shape of the
% distribution.  If the number of bins was greater than 20, the bars do not
% always touch.

% Part b
mean(mother_heights)
median(mother_heights)
std(mother_heights)
iqr(mother_heights)
% Yes, the center of the sample seems well-defined because the mean and
% median are both approximately 64.

% Part c
boxplot(mother_heights)
snapnow
plot(ecdf(mother_heights))
snapnow
qqplot(mother_heights)
snapnow
% Yes, it appears that this sample is approximately normal considering the
% boxplot (placement of mean relative to Q1/Q3, proper spacing), eCDF plot
% (matches well with normal eCDF), QQ plot (linearity mostly preserved),
% and histogram (decently strong bell-shape).  The parameter for mean would
% simply be 64.04 and the parameter for st. dev. would be 2.53.

% Part d
smokers = birth(birth(:,7)==1,:);
smokers_heights = smokers(:,5);
smokers_heights = smokers_heights(smokers_heights<99);

nonsmokers = birth(birth(:,7)==0,:);
nonsmokers_heights = nonsmokers(:,5);
nonsmokers_heights = nonsmokers_heights(nonsmokers_heights<99);

boxplot(smokers_heights)
snapnow
boxplot(nonsmokers_heights)
snapnow
% Comparing the boxplots, the heights of smokers vs. nonsmokers in this
% case are very similar.  The means and medians appear to be the same.
% Thus, we cannot say that one of the two groups has definitively higher
% average heights.
```

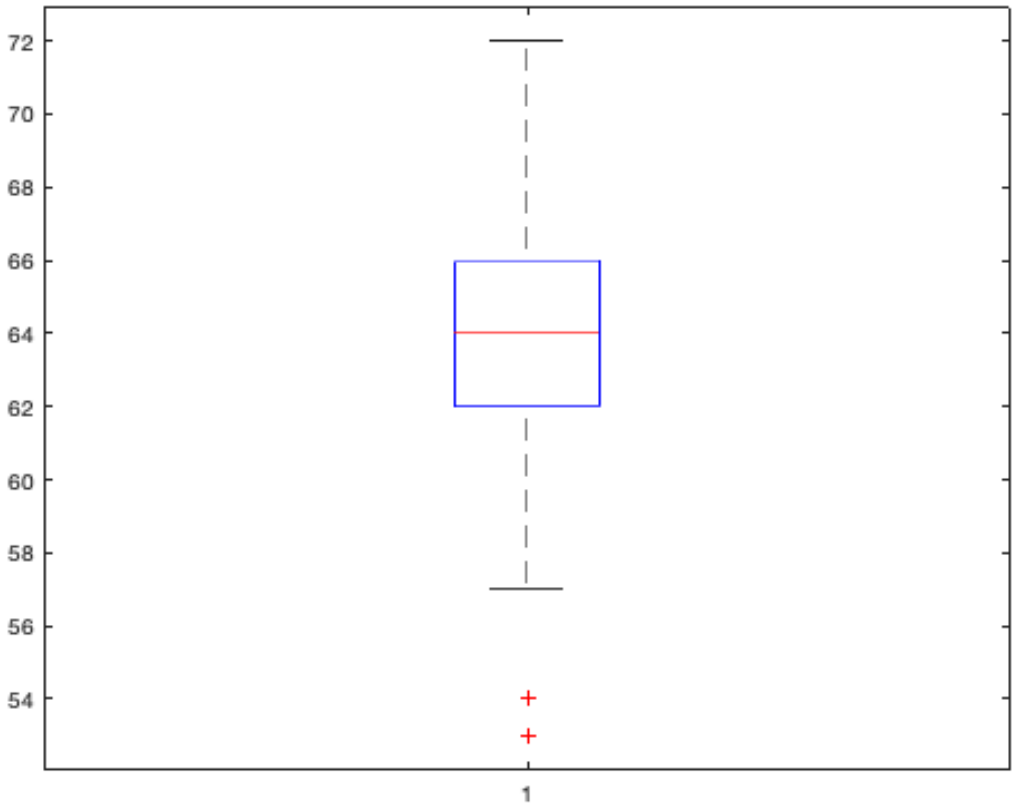


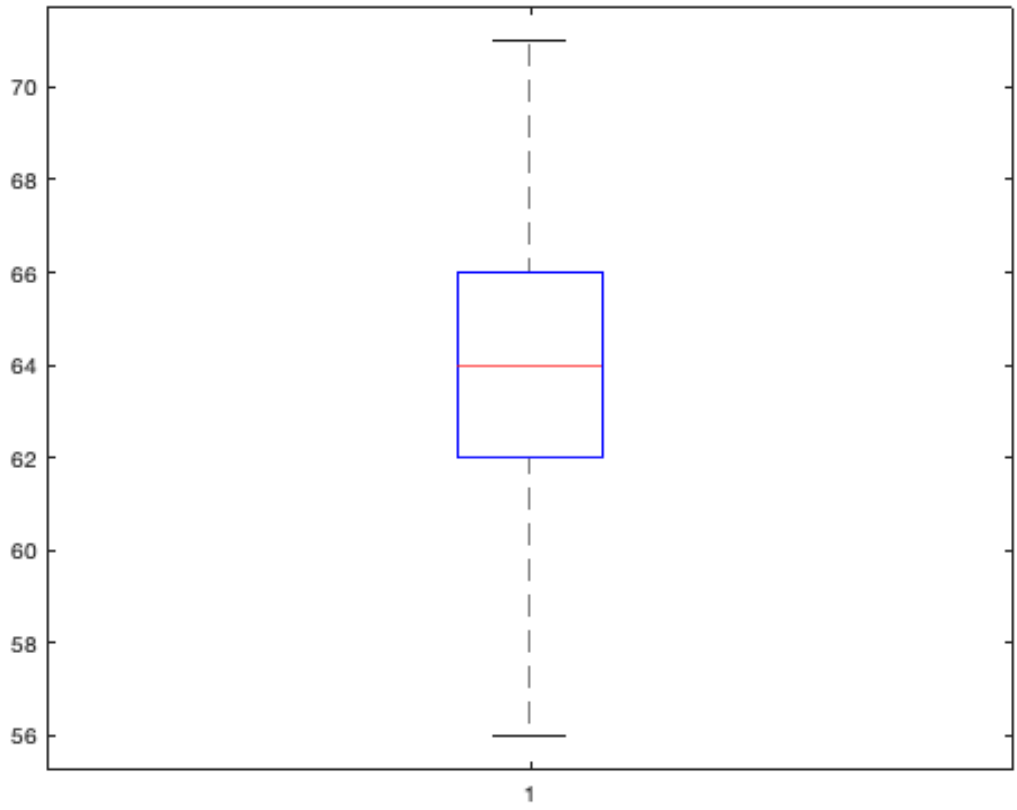

```
ans =

   64.0478

ans =

    64

ans =

    2.5334

ans =

     4
```


72
70
68
66
64
62
60
58
56
54
1


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20


QQ Plot of Sample Data versus Standard Normal
Quantiles of Input Sample
Standard Normal Quantiles
75
70
65
60
55
50
-4
-3
-2
-1
0
1
2
3
4


72
70
68
66
64
62
60
58
56
54
1



Published with MATLAB® R2021b