STATS-S631

Assignment 1

John Koo

Problem 1

$S = \{\text{set of all UN members}\}$

$X : S \to \mathbb{R}$, maps a female life expectancy from each country

X is a random variable since $\forall r \in \mathbb{R}$ there is a probability that $X(s) = r$ for some $s \in S$.

Problem 2

a. $\{s \in S : X(s) \leq 80\}$

b. $\{s \in S : X(s) = 75\}$

c. $\{s \in S : X(s) \in [65, 70]\}$

Problem 3

a. $P(S \leq 80)$

b. $P(S \leq 75) - P(S \leq 74)$

c. $P(S \leq 70) - P(S \leq 64)$

Problem 4

First, load the data as a variable

```
un.df <- alr4::UN11
```

Then compute the probabilities.

```
part.a <- nrow(un.df[un.df$lifeExpF <= 80, ]) / nrow(un.df)

part.b <- nrow(un.df[un.df$lifeExpF == 75, ]) / nrow(un.df)

part.c <-
  nrow(un.df[un.df$lifeExpF <= 70 & un.df$lifeExpF >= 65, ]) / nrow(un.df)
```

(Results rounded to 3 decimal places)

a. 0.769

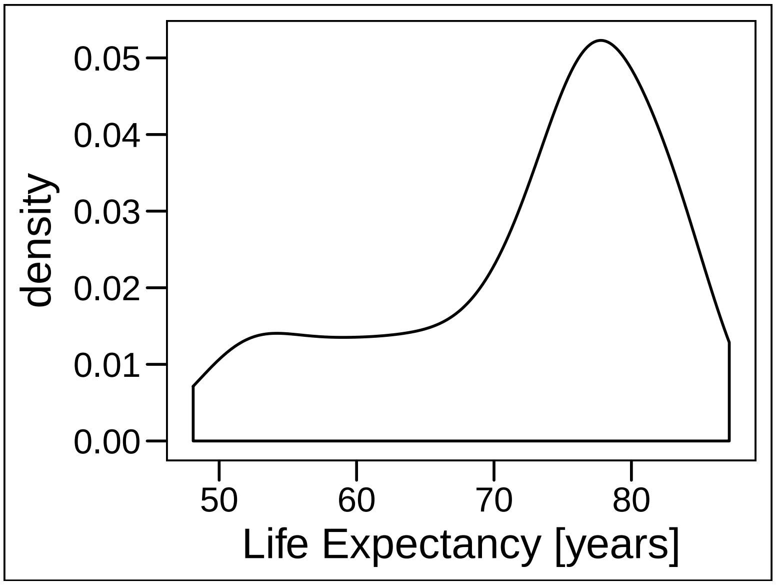
b. 0.005

c. 0.075

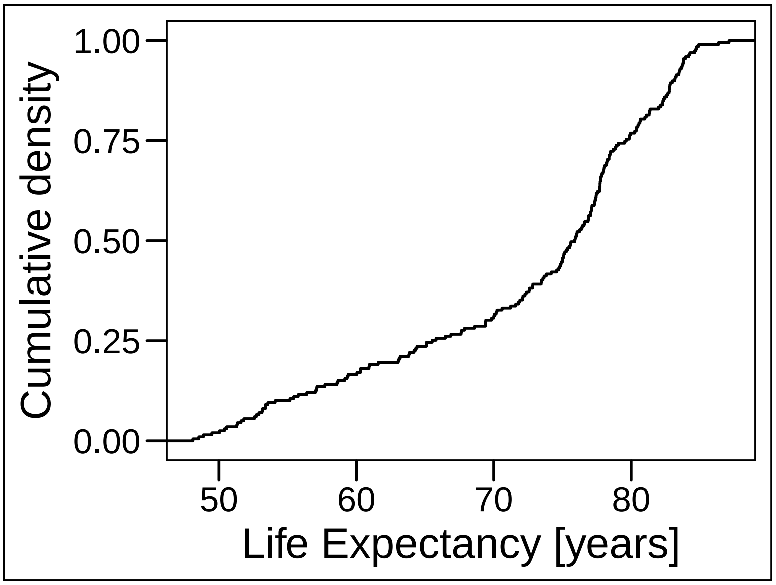
Problem 5

```
import::from(ggplot2,
             ggplot, theme_set, aes, geom_density, labs, stat_ecdf)
theme_set(ggthemes::theme_base())

ggplot(un.df) +
  geom_density(aes(x = lifeExpF)) +
  labs(x = 'Life Expectancy [years]')
```



```
ggplot(un.df) +
  stat_ecdf(aes(x = lifeExpF)) +
  labs(x = 'Life Expectancy [years]',
       y = 'Cumulative density')
```



The distribution does not appear normal. It's not symmetric and there is a clear skew to the left. It might be possible to approximate it as the sum of two normal distributions. We can verify this by computing the skewness of the distribution:

```
print(moments::skewness(un.df$lifeExpF))
```

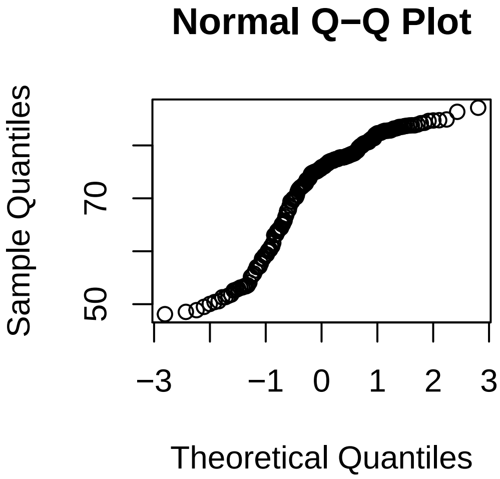
```
[1] -0.859599
```

Since the skewness is much closer to -1 than to 0, we can say that a symmetric distribution, such as the normal distribution, is not a good approximation for these data.

Although it's pretty clear at this point that the population cannot be approximated well with a normal distribution, if it weren't clear, we could do a few quick tests.

One such test is to check the quantile-quantile plot. Since the points do not fall in a straight line, we can say that a normal distribution cannot approximate the data:

```
qqnorm(un.df$lifeExpF)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles