Creating Graphs for the Data

(Data Management and Visualization Week 4 Assignment)

Expected Activities

STEP 1: Create graphs of your variables one at a time (univariate graphs).

Examine both their center and spread.

STEP 2: Create a graph showing the association between your explanatory and response variables (bivariate graph).

Output should be interpretable (i.e. organized and labeled).

SAS Program

```
LIBNAME mydata "/courses/d1406ae5ba27fe300 " ACCESS=readonly;

DATA new;
	SET mydata.gapminder;
	KEEP country lifeexpectancy urbanrate femaleemployrate urban le fer;
	LABEL lifeexpectancy="Life Expectancy";
	LABEL urbanrate="Urbanisation Rate";
	LABEL femaleemployrate="Female Employment Rate";
	LABEL urban="Urbanisation Groups";
	LABEL le="Life Expectancy Groups";
	LABEL fer="Female Employment Rate Groups";

	/* Data Preparation or Management for variable urbanrate */
	IF urbanrate < 25 THEN
		urban="UR Group 1";
	IF urbanrate >=25 AND urbanrate < 50 THEN
		urban="UR Group 2";
	IF urbanrate >=50 AND urbanrate < 75 THEN
		urban="UR Group 3";
	IF urbanrate >=75 THEN
		urban="UR Group 4";

	/* Data Preparation or Management for variable lifeexpectancy */
	IF lifeexpectancy <  40 THEN
		le="LE Group 1";
	IF lifeexpectancy >=40 AND lifeexpectancy < 50 THEN
		le="LE Group 2";
	IF lifeexpectancy >=50 AND lifeexpectancy < 60 THEN
		le="LE Group 3";
	IF lifeexpectancy >=60 AND lifeexpectancy < 70 THEN
		le="LE Group 4";
	IF lifeexpectancy >=70 THEN
		le="LE Group 5";
```

```
        /* Data Preparation or Management for variable femaleemployrate */
        IF femaleemployrate <  20 THEN
                fer="FER Group 1";
        IF femaleemployrate >=20 AND femaleemployrate < 30 THEN
                fer="FER Group 2";
        IF femaleemployrate >=30 AND femaleemployrate < 40 THEN
                fer="FER Group 3";
        IF femaleemployrate >=40 AND femaleemployrate < 50 THEN
                fer="FER Group 4";
        IF femaleemployrate >=50 AND femaleemployrate < 60 THEN
                fer="FER Group 5";
        IF femaleemployrate >=60 THEN
                fer="FER Group 6";

PROC SORT;
        BY COUNTRY;

PROC GCHART;
        VBAR urbanrate/ TYPE=PCT;
        Title 'Univariate Bar Graph - Frequency Distribution';
        Title2 'Urbanisation Rate';

PROC GCHART;
        VBAR urban/DISCRETE TYPE=PCT;
        Title 'Univariate Bar Graph - Categorical Frequency Distribution';
        Title2 'Urbanisation Groups';

PROC GCHART;
        VBAR lifeexpectancy/ TYPE=PCT;
        Title 'Univariate Bar Graph - Frequency Distribution';
        Title2 'Life Expectancy';

PROC GCHART;
        VBAR le/DISCRETE TYPE=PCT;
        Title 'Univariate Bar Graph - Categorical Frequency Distribution';
        Title2 'Life Expectancy Groups';

PROC GCHART;
        VBAR femaleemployrate/ TYPE=PCT;
        Title 'Univariate Bar Graph - Frequency Distribution';
        Title2 'Female Employment Rate';

PROC GCHART;
        VBAR fer/DISCRETE TYPE=PCT;
        Title 'Univariate Bar Graph - Categorical Frequency Distribution';
        Title2 'Female Employment Rate Groups';

PROC UNIVARIATE;
        VAR urbanrate lifeexpectancy;
        Title 'Univariate Analysis';
        Title2 'Urbanisation Rate Vs Life Expectancy';

PROC GPLOT;
        PLOT lifeexpectancy*urbanrate;
        Title 'Bivariate Scatter Plot';
```

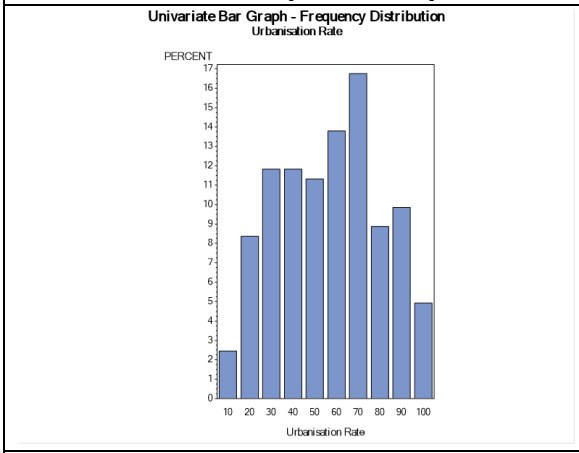
```
        Title2 'Urbanisation Rate Vs Life Expectancy';

PROC GCHART;
        VBAR urban/DISCRETE TYPE=mean SUMVAR=lifeexpectancy;
        Title 'Bivariate Bar Graph';
        Title2 'Urbanisation Groups Vs Life Expectancy';

PROC UNIVARIATE;
        VAR urbanrate femaleemployrate;
        Title 'Univariate Analysis';
        Title2 'Urbanisation Rate Vs Female Employment Rate';

PROC GPLOT;
        PLOT femaleemployrate*urbanrate;
        Title 'Bivariate Scatter Plot';
        Title2 'Urbanisation Rate Vs Female Employment Rate';

PROC GCHART;
        VBAR urban/DISCRETE TYPE=mean SUMVAR=femaleemployrate;
        Title 'Bivariate Bar Graph';
        Title2 'Urbanisation Groups Vs Female Employment Rate';

RUN;
```

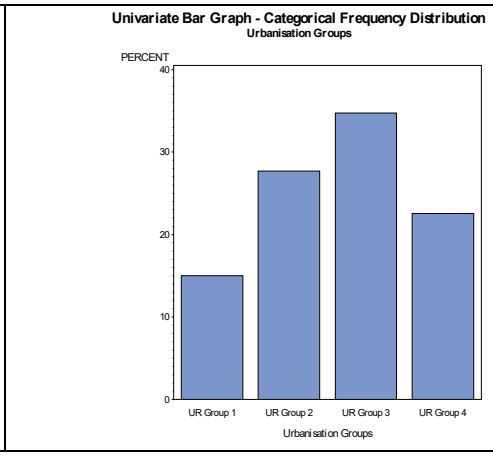
Output – Univariate & Bivariate Graphs

Univariate graphs are created for the following three variables from the Gapminder dataset.

- Urbanisation Rate (urbanrate)
- Life Expectancy (lifeexpectancy)
- Female Employment Rate (femaleemployrate)

Urbanisation Rate (urbanrate)

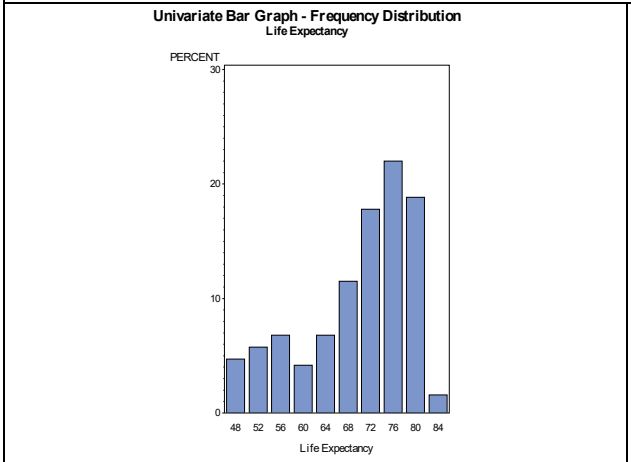


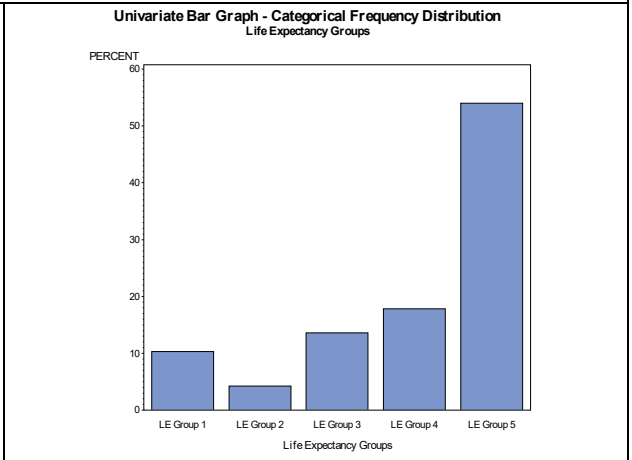


The average urbanization rate is 56.77% and the standard deviation (spread) is 23.8449326. This means that the differences between results are quite high.

Urbanization Rate is classified into four different groups - UR Group 1 (urbanrate < 25), UR Group 2 (25 <= urbanrate < 50), UR Group 3 (50 <= urbanrate < 75), and UR Group 4

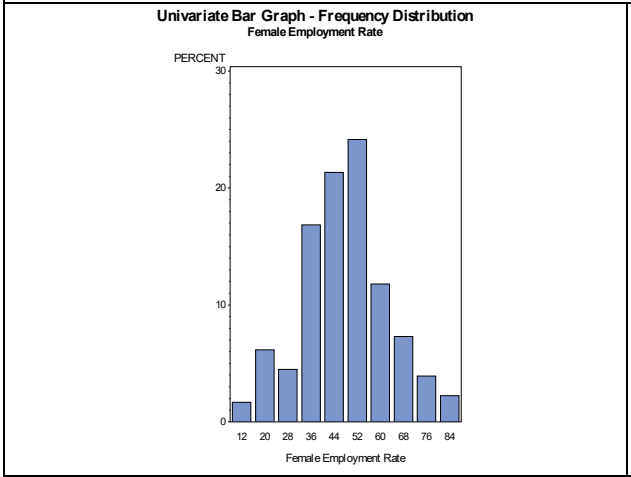
(urbanrate >= 75). Majority of records are falling under UR Group 3 (~ 35%) followed by UR Group 2 (~28%). The lowest frequency is observed in UR Group 1(~15%).

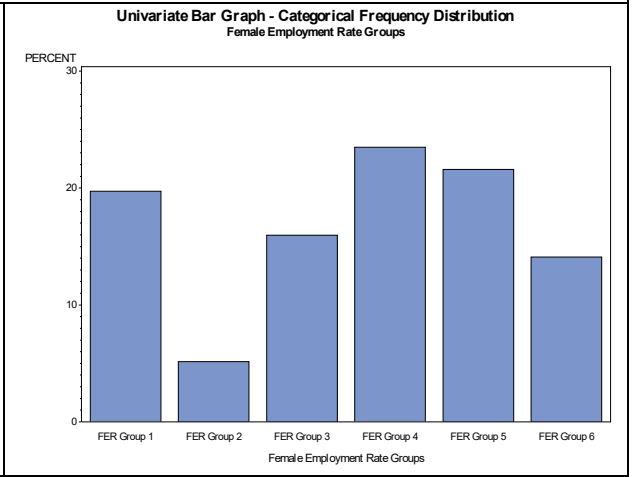




The average life expectancy is 69.75 and the standard deviation (spread) is 9.70862054. This means that the differences between results are not too high.

Life Expectancy is classified into five different groups - LE Group 1 (lifeexpectancy < 40), LE Group 2 (40 <= lifeexpectancy < 50), LE Group 3 (50 <= lifeexpectancy < 60), LE Group 4 (60 <= lifeexpectancy < 70), and LE Group 5 (lifeexpectancy >= 70). Majority of records are falling under LE Group 5 (~ 54%) and the lowest frequency is observed in LE Group 2 (~4%)..



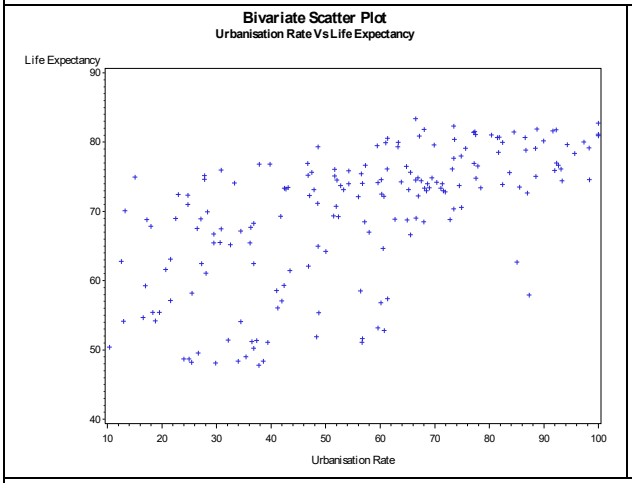


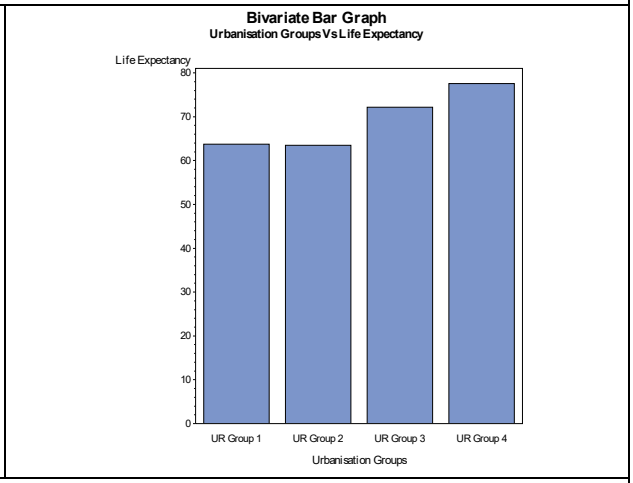
The average female exployment rate is 47.55% and the standard deviation (spread) is 14.62574. This means that the differences between results are significant.

Female Employment Rate is classified into six different groups - FER Group 1 (femaleemployrate < 20), FER Group 2 (20 <= femaleemployrate < 30), FER Group 3 (30 <= femaleemployrate < 40), FER Group 4 (40 <= femaleemployrate < 50), FER Group 5 (50 <= femaleemployrate < 60), and FER Group 6 (femaleemployrate >= 60). Majority of the data is falling under FER Group 4 (~ 23%) followed by FER Group 5 (~22%). The lowest frequency is observed in FER Group 2 (~5%).

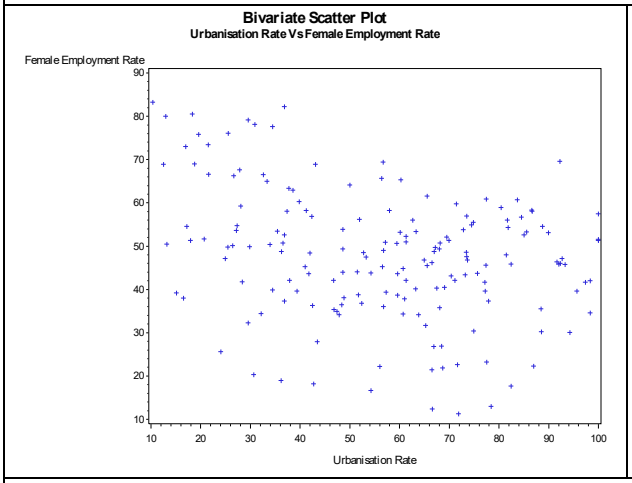
Bivariate graphs are created for the following two scenarios.

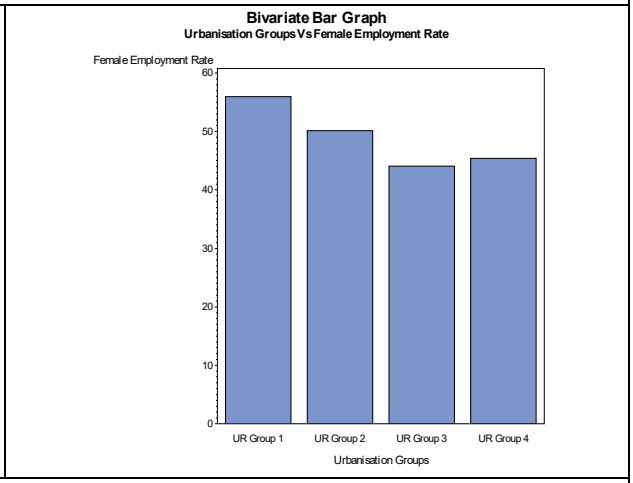
- Urbanisation Rate Vs Life Expectancy (urbanrate Vs lifeexpectancy)
- Urbanisation Rate Vs Female Employment Rate (urbanrate Vs femaleemployrate)





The graphs show the relationship between Life Expectancy and Urbanization rate. We can see a trend that there is more life expectancy with the higher urbanization rate.





The graphs show the relationship between Urbanisation Rate and Female Employment Rate. We can see a trend that there is a slight decrease in female employment rate with the higher urbanization.