

Types of Graphs

1. Stem and Leaf
2. Histogram (Frequency Bar Chart)
3. Cumulative Frequency Plot (ECDF)
4. Boxplots
5. Scatterplots (not discussed here)

Graphs are used to visualize, inspect, and present data.

Graphs

DATA: Male Life Expectancy by Country (from 1993 Demographic Yearbook)

. list

	country	life	rlife
1.	Canada	73.02	73
2.	Coata Rica	72.89	73
3.	Cuba	72.74	73
4.	United States	72	72
5.	Jamaica	71.41	71
6.	Bermuda	70.23	70
7.	Panama	69.78	70
8.	Bahamas	68.32	68
9.	Aruba	68.3	68
10.	Bardados	67.15	67
11.	Nicaragua	64.8	65
12.	Mexico	62.1	62
13.	Greenland	60.4	60
14.	Haiti	54.95	55
15.	El Salvador	50.74	51
16.	Netherlands	71.13	71
17.	Belize	69.95	70
18.	Puerto Rico	69.6	70
19.	Saint Lucie	68	68
20.	Dominican Rep	67.63	68
21.	Martinque	67	67
22.	Trinidad & Tobago	66.88	67
23.	Guadeloupe	66.4	66
24.	Saint Kitts & Nevis	65.87	66
25.	Honduras	65.43	65
26.	Guatemala	55.11	55

Graphs

Stem and Leaf Plots

- Aids in sorting or ordering data
- Pick the number of stems carefully

```
. stem rlife
```

Stem-and-leaf plot for rlife

5*		1
5.		55
6*		02
6.		55667778888
7*		0000112333

```
. stem rlife, lines(5)
```

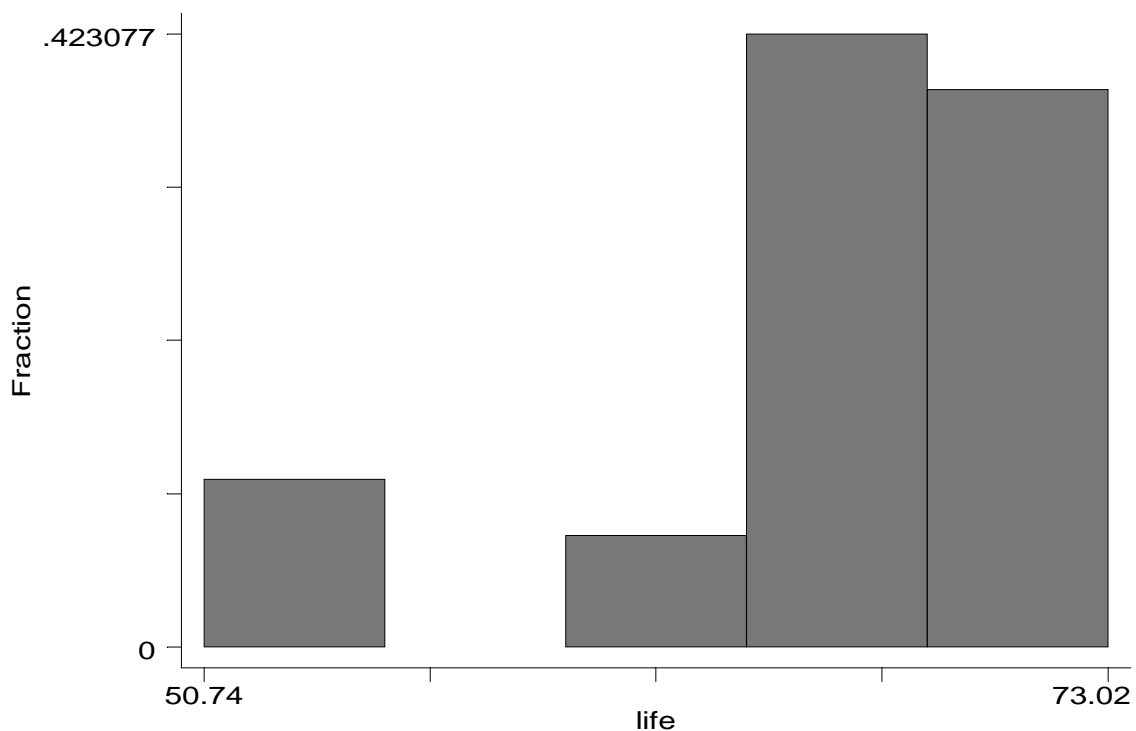
Stem-and-leaf plot for rlife

5*		1
5t		
5f		55
5s		
5.		
6*		0
6t		2
6f		55
6s		66777
6.		8888
7*		000011
7t		2333

Histograms (Frequency Bar Chart)

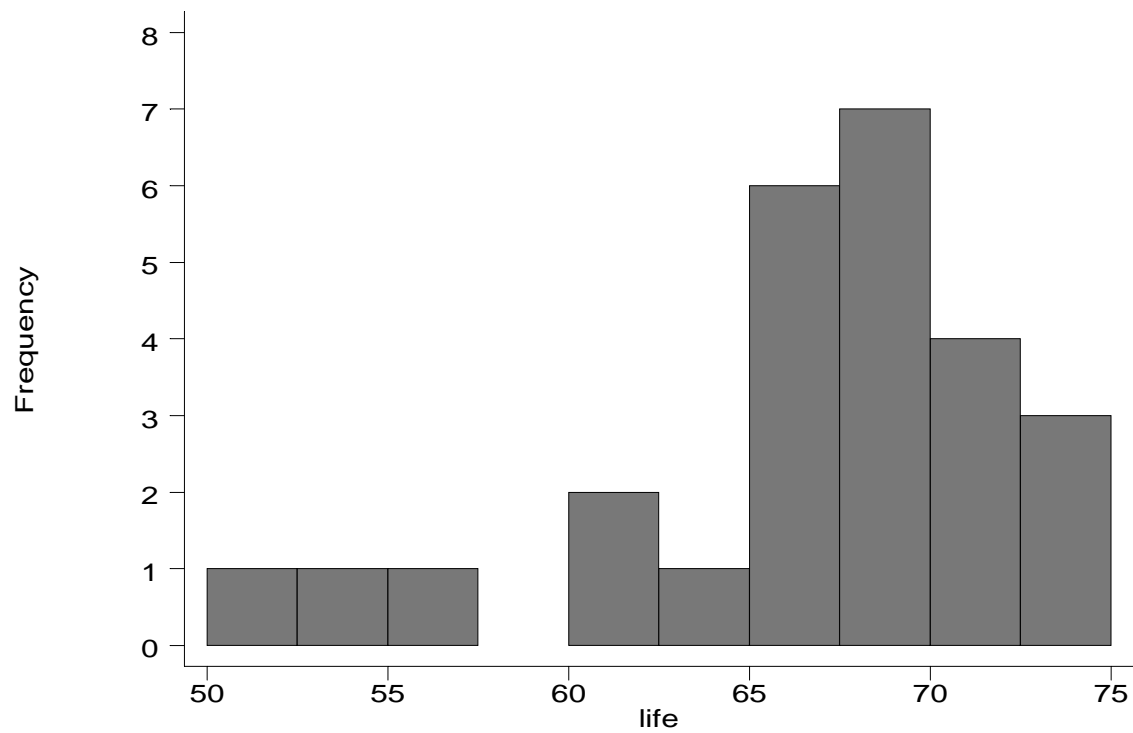
- Plot of group type by number of observations in that group
- Width of groups, should be the same!
- Again, the number of groups (called bins) is important.

```
. hist life
```



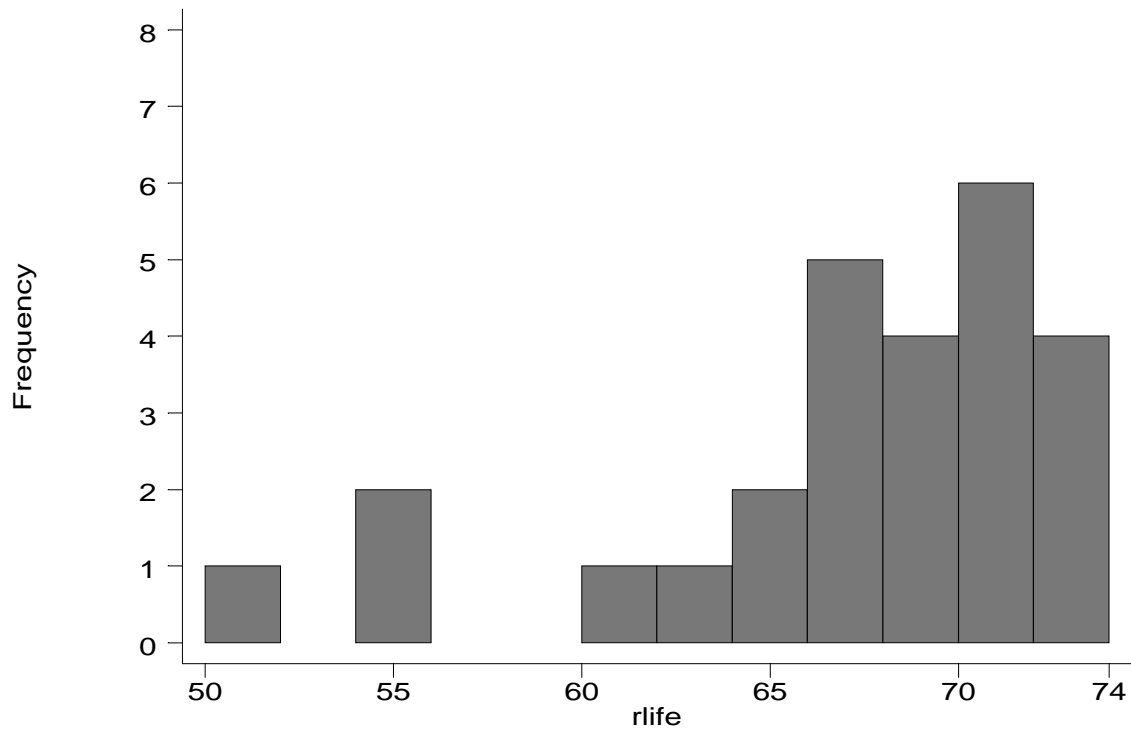
Graphs

```
. graph7 life, hi f bin(10) ylabel(0,1,2,3,4,5,6,7,8)  
                        xlabel(50,55,60,65,70,7)
```



Graphs

```
. graph7 rlife, hi f bin(12) ylabel(0,1,2,3,4,5,6,7,8)
      xlabel(50,55,60,65,70,74)
```



```
. stem rlife, width(2)
```

Stem-and-leaf plot for rlife

5*	1
5t	
5f	55
5s	
5.	
6*	0
6t	2
6f	55
6s	66777
6.	8888
7*	000011
7t	2333

Graphs

Cumulative Frequency Plot (ECDF)

- (Empirical Cumulative Distribution Function)
- Plot the cumulative frequency versus variable
- Helps to visualize the centiles

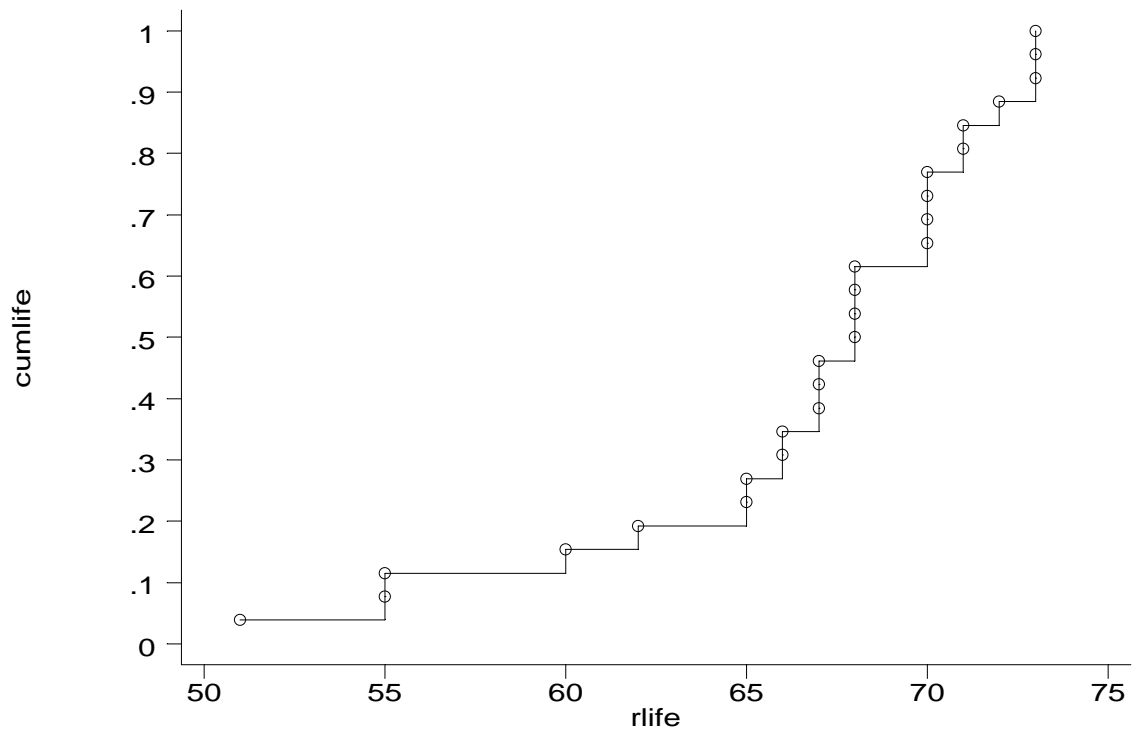
```
. cumul rlife, gen(cumlife)
```

```
. list
```

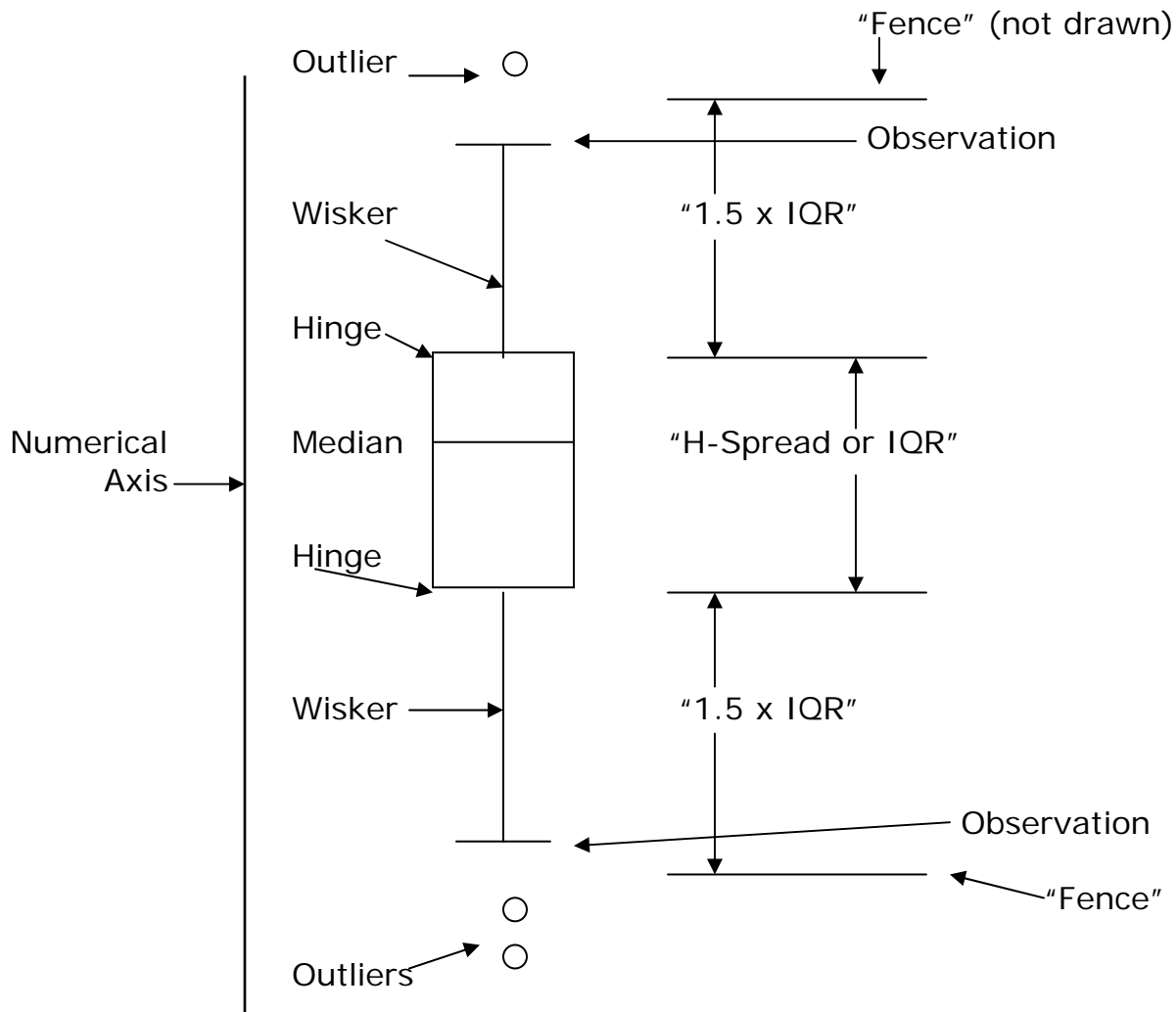
	country	life	rlife	cumlife
1.	El Salvador	50.74	51	.0384615
2.	Haiti	54.95	55	.0769231
3.	Guatemala	55.11	55	.1153846
4.	Greenland	60.4	60	.1538462
5.	Mexico	62.1	62	.1923077
6.	Honduras	65.43	65	.2307692
7.	Nicaragua	64.8	65	.2692308
8.	Saint Kitts & Nevis	65.87	66	.3076923
9.	Guadeloupe	66.4	66	.3461539
10.	Martinique	67	67	.3846154
11.	Bardados	67.15	67	.4230769
12.	Trinidad & Tobago	66.88	67	.4615385
13.	Aruba	68.3	68	.5
14.	Saint Lucie	68	68	.5384616
15.	Bahamas	68.32	68	.5769231
16.	Dominican Rep	67.63	68	.6153846
17.	Puerto Rico	69.6	70	.6538461
18.	Belize	69.95	70	.6923077
19.	Panama	69.78	70	.7307692
20.	Bermuda	70.23	70	.7692308
21.	Jamaica	71.41	71	.8076923
22.	Netherlands	71.13	71	.8461539
23.	United States	72	72	.8846154
24.	Coata Rica	72.89	73	.9230769
25.	Cuba	72.74	73	.9615384
26.	Canada	73.02	73	1

Graphs

```
. graph7 cumlife rlife,connect(J)
      ylabel(0,.1,.2,.3,.4,.5,.6,.7,.8,.9,1)
      xlabel(50,55,60,65,70,75)
```



Anatomy of a Boxplot

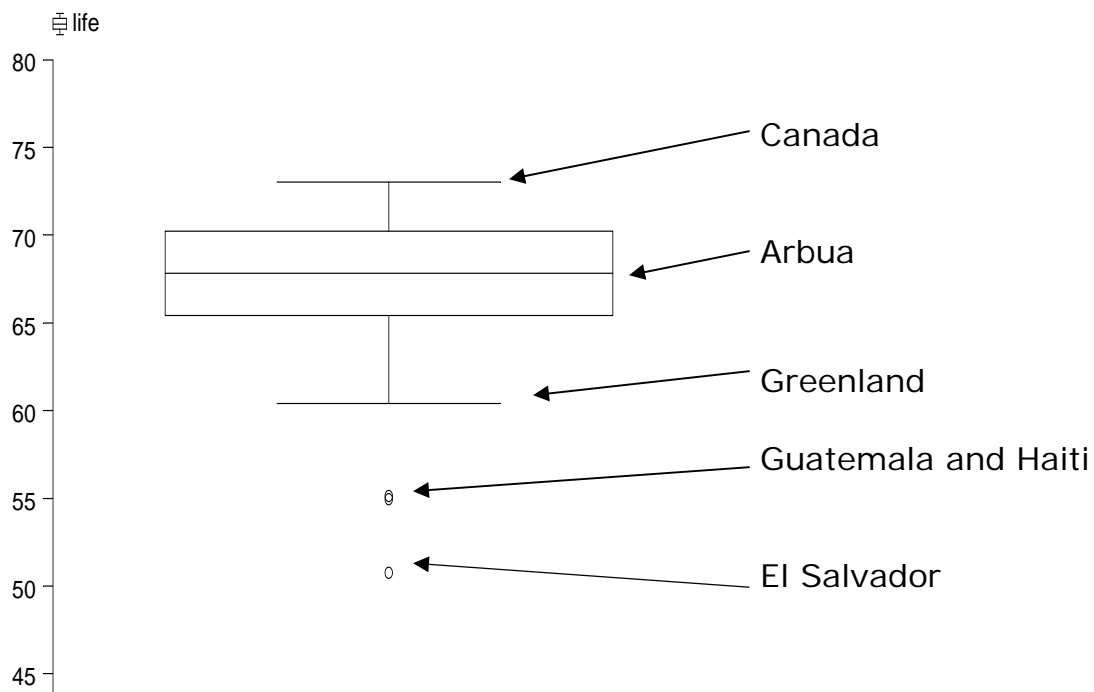


- Median is the 50th percentile
- Hinges are the 25th and 75th percentile
- IQR=difference between 75th and 25th centile
- Fences are not drawn on boxplot
- Whiskers are drawn to the closest observation not beyond fence
- Outliers are observations beyond the fence

Graphs

Example: Boxplot of Male Life Expectancy

```
. graph7 life, box ylab(45,50,55,60,65,70,75,80)
```



- Yuck to Stata's width for single boxplots...
- See handout for more examples!