

STAT UN1201 – Chapter 1

Prof. Joyce Robbins

Waitlist

1. The waitlist moves in order as places open up.
2. Course materials are available here during change of program period: <http://github.com/jtr13/1201>
3. It is strongly advised to keep up with the material if you are trying to get in the class.

EVERYONE: Once you've made a decision not to take the class, please be considerate and drop it from your schedule.

Very brief course outline

- Descriptive statistics (Chapter 1)
- Probability (Chapters 2 - 5)
- Inferential statistics (Chapters 6 - 9, 12 - 13)

Descriptive statistics

- numerical summaries: mean, median, etc.
- graphical summaries

Anscombe's Quartet

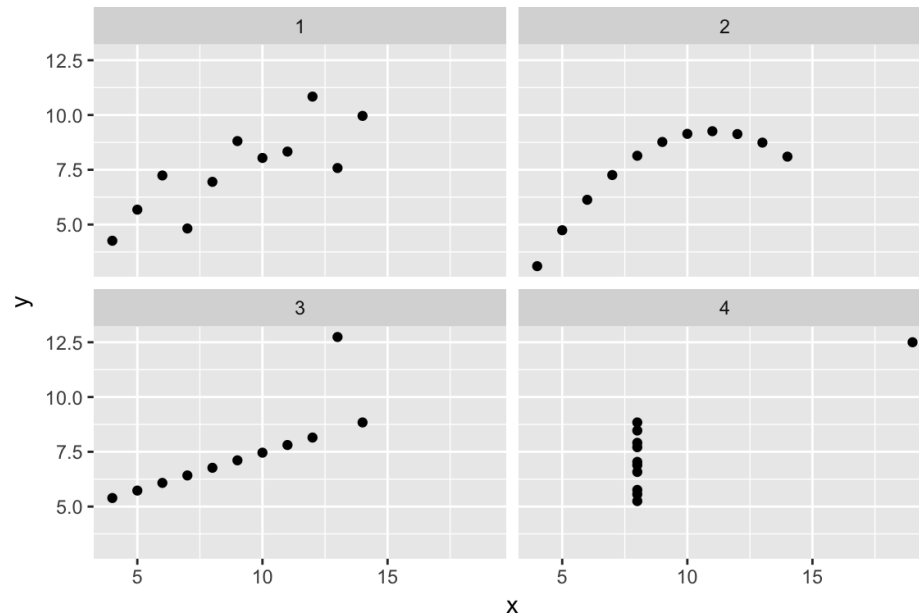
x1	x2	x3	x4	y1	y2	y3	y4
10	10	10	8	8.04	9.14	7.46	6.58
8	8	8	8	6.95	8.14	6.77	5.76
13	13	13	8	7.58	8.74	12.74	7.71
9	9	9	8	8.81	8.77	7.11	8.84
11	11	11	8	8.33	9.26	7.81	8.47
14	14	14	8	9.96	8.10	8.84	7.04
6	6	6	8	7.24	6.13	6.08	5.25
4	4	4	19	4.26	3.10	5.39	12.50
12	12	12	8	10.84	9.13	8.15	5.56
7	7	7	8	4.82	7.26	6.42	7.91
5	5	5	8	5.68	4.74	5.73	6.89

Anscombe's Quartet, numerical summaries

For each of the four data sets:

- Number of observations (n) = 11
- Mean of the x 's (\bar{x}) = 9.0
- Mean of the y 's (\bar{y}) = 7.6
- Equation of the regression line: $y = 3 + 0.5x$

Anscombe's Quartet, graphical analysis



Collecting data

CITYREALTY
NEW YORK CITY REAL ESTATE

SEARCH NYC APARTMENTS **MARKET REPORTS** **GE**

MANHATTAN **BROOKLYN** **QUEENS & BRONX** **SEARCH BY MAP** **ADVANCED SEARCH**

<input type="checkbox"/> UPPER EAST SIDE <input type="checkbox"/> Park/Fifth Ave. to 79th St. <input type="checkbox"/> Lenox Hill <input type="checkbox"/> Carnegie Hill <input type="checkbox"/> Yorkville <input type="checkbox"/> Roosevelt Island	<input type="checkbox"/> MIDTOWN <input type="checkbox"/> Beekman/Sutton Place <input type="checkbox"/> Midtown East <input type="checkbox"/> Murray Hill <input type="checkbox"/> Turtle Bay/United Nations <input type="checkbox"/> Midtown West	<input type="checkbox"/> DOWNTOWN <input type="checkbox"/> Chelsea <input type="checkbox"/> Gramercy Park <input type="checkbox"/> Greenwich Village <input type="checkbox"/> East Village <input type="checkbox"/> West Village <input type="checkbox"/> Lower East Side <input type="checkbox"/> NoLiTa/Little Italy <input type="checkbox"/> NoHo <input type="checkbox"/> SoHo <input type="checkbox"/> Tribeca <input type="checkbox"/> Flatiron/Union Square <input type="checkbox"/> Stuyvesant Town / PCV	<input type="checkbox"/> UPPER MANHATTAN <input type="checkbox"/> Harlem <input type="checkbox"/> East Harlem <input type="checkbox"/> Washington Heights <input type="checkbox"/> Inwood <input type="checkbox"/> Hamilton Heights
<input checked="" type="checkbox"/> UPPER WEST SIDE <input type="checkbox"/> Broadway Corridor <input type="checkbox"/> Central Park West <input type="checkbox"/> Lincoln Center <input checked="" type="checkbox"/> Morningside Heights <input type="checkbox"/> Riverside Dr./West End Ave.	<input type="checkbox"/> FINANCIAL DISTRICT/BPC <input type="checkbox"/> Battery Park City <input type="checkbox"/> Financial District		<input checked="" type="checkbox"/> ALL MANHATTAN

SALES **RENTALS** **PRICE RANGE:** \$100K - \$50M+ **BEDROOMS** Select (1) **TYPE** Select **VIEW LISTINGS →**

Collecting data

HIDE ALL PHOTOS SHOW ALL PHOTOS

ADDRESS	TYPE	NEIGHBORHOOD	BEDS / BATH / SIZE	PRICE	PRICE/FT ²	LISTED
395 Riverside Drive, #4E	CO-OP	Morningside Heights	1 bd / 1 ba	\$799K	-	Feb 22, 2016
390 Riverside Drive, #4G	CO-OP	Morningside Heights	1 bd / 1 ba / 800 ft ²	\$745K	\$931	Jul 19, 2016
380 Riverside Drive, #8L	CO-OP	Morningside Heights	1 bd / 1 ba / 700 ft ²	\$725K	\$1,036	Sep 06, 2016
535 West 110th Street, #6B	CO-OP	Morningside Heights	1 bd / 1 ba / 750 ft ²	\$725K	\$967	Sep 07, 2016
535 West 110th Street, #6A	CO-OP	Morningside Heights	1 bd / 1 ba / 750 ft ²	\$699K (-4% ▼)	\$933	Apr 26, 2016
395 Riverside Drive, #2D	CO-OP	Morningside Heights	1 bd / 1 ba / 900 ft ²	\$699K	\$777	Mar 03, 2016
528 West 111th Street, #24	CO-OP	Morningside Heights	1 bd / 1 ba	\$675K (-8.9% ▼)	-	Apr 15, 2015
545 West 111th Street, #9N	CO-OP	Morningside Heights	1 bd / 1 ba / 650 ft ²	\$665K (-9.5% ▼)	\$1,023	Mar 30, 2016
380 Riverside Drive, #6B	CO-OP	Morningside Heights	1 bd / 1 ba	\$599K	-	Oct 07, 2015
80 La Salle Street, #20C	CO-OP	Morningside Heights	1 bd / 1 ba	\$545K	-	May 19, 2016
609 West 114th Street, #36	CO-OP	Morningside Heights	1 bd / 1 ba / 600 ft ²	\$535K	\$892	Jun 22, 2016
80 La Salle Street, #21B	CO-OP	Morningside Heights	1 bd / 1 ba / 725 ft ²	\$535K	\$738	May 04, 2016
80 La Salle Street, #14D	CO-OP	Morningside Heights	1 bd / 1 ba	\$529K	-	Apr 14, 2016
175 Claremont Avenue, #3	CO-OP	Morningside Heights	1 bd / 1 ba	\$499K (-9.3% ▼)	-	Mar 30, 2016
3117 Broadway, #56	CO-OP	Morningside Heights	1 bd / 1 ba / 500 ft ²	\$450K	\$900	Jun 28, 2016

FEATURED NEW DEVELOPMENTS



510 West 123rd Street, #38	CO-OP	Morningside Heights	1 bd / 1 ba / 700 ft ²	\$450K	\$643	Jul 02, 2016
3115 Broadway, #43	CO-OP	Morningside Heights	1 bd / 1 ba / 650 ft ²	\$425K	\$654	Apr 18, 2016
3115 Broadway, #42	CO-OP	Morningside Heights	1 bd / 1 ba / 650 ft ²	\$379K (-5% ▼)	\$583	Mar 28, 2016

Cleaning / organizing data

	A	B	C	D	E	F	G	H	
1		ADDRESS	TYPE	NEIGHBORHOOD	BEDS / BATH	PRICE	PRICE/FT2	LISTED	
2	+								
3	3115 Broadw	CO-OP							
4	Morningside	1 bd / 1 ba /	\$379K	\$583	28-Mar-16				
5	+								
6	3115 Broadw	CO-OP							
7	Morningside	1 bd / 1 ba /	\$425K	\$654	18-Apr-16				
8	+								
9	510 West 12	CO-OP							
10	Morningside	1 bd / 1 ba /	\$450K	\$643	2-Jul-16				
11	+								
12	3117 Broadw	CO-OP							
13	Morningside	1 bd / 1 ba /	\$450K	\$900	28-Jun-16				
14	+								
15	175 Claremo	CO-OP							
16	Morningside	1 bd / 1 ba	\$499K	-	30-Mar-16				
17	+								
18	80 La Salle St	CO-OP							
19	Morningside	1 bd / 1 ba	\$529K	-	14-Apr-16				

Cleaning / organizing data

	A	B	C	D
1	Address		Price	IntPrice
2	3115 Broadway, #42	1 bd / 1 ba / 650 ft2	\$379K	379
3	3115 Broadway, #43	1 bd / 1 ba / 650 ft2	\$425K	425
4	3117 Broadway, #56	1 bd / 1 ba / 500 ft2	\$450K	450
5	510 West 123rd Street, #38	1 bd / 1 ba / 700 ft2	\$450K	450
6	175 Claremont Avenue, #3	1 bd / 1 ba	\$499K	499
7	80 La Salle Street, #14D	1 bd / 1 ba	\$529K	529
8	609 West 114th Street, #36	1 bd / 1 ba / 600 ft2	\$535K	535
9	80 La Salle Street, #21B	1 bd / 1 ba / 725 ft2	\$535K	535
10	80 La Salle Street, #20C	1 bd / 1 ba	\$545K	545
11	380 Riverside Drive, #6B	1 bd / 1 ba	\$599K	599
12	545 West 111th Street, #9N	1 bd / 1 ba / 650 ft2	\$665K	665
13	528 West 111th Street, #24	1 bd / 1 ba	\$675K	675
14	535 West 110th Street, #6A	1 bd / 1 ba / 750 ft2	\$699K	699
15	395 Riverside Drive, #2D	1 bd / 1 ba / 900 ft2	\$699K	699
16	380 Riverside Drive, #8L	1 bd / 1 ba / 700 ft2	\$725K	725
17	535 West 110th Street, #6B	1 bd / 1 ba / 750 ft2	\$725K	725
18	390 Riverside Drive, #4G	1 bd / 1 ba / 800 ft2	\$745K	745
19	395 Riverside Drive, #4E	1 bd / 1 ba	\$799K	799
20				
21				

Types of data

- Quantitative
 - *Discrete (finite or can be listed in an infinite sequence)*
 - *Continuous (entire continuum of possible values)*
- Qualitative (categorical)

Visualizing one-dimensional continuous data

- stem-and-leaf plot
- histogram
 - *frequency*
 - *relative frequency*
 - *density*
- boxplot

Stem-and-leaf plot

Prices:

```
## [1] 379 425 450 450 499 529 535 535 545 599 665  
## [12] 675 699 699 725 725 745 799
```

Prices, rounded to 2 significant digits:

```
## [1] 380 430 450 450 500 530 540 540 550 600 670  
## [12] 680 700 700 730 730 750 800
```

Stem-and-leaf plot

Prices:

```
## [1] 379 425 450 450 499 529 535 535 545 599 665
## [12] 675 699 699 725 725 745 799
```

Prices, rounded to 2 significant digits:

```
## [1] 380 430 450 450 500 530 540 540 550 600 670
## [12] 680 700 700 730 730 750 800
```

```
##
## The decimal point is 2 digit(s) to the right of the |
##
## 3 | 8
## 4 | 355
## 5 | 03445
## 6 | 078
## 7 | 00335
## 8 | 0
```

Stem and leaf plot

Total Fertility Rate (average births per woman)

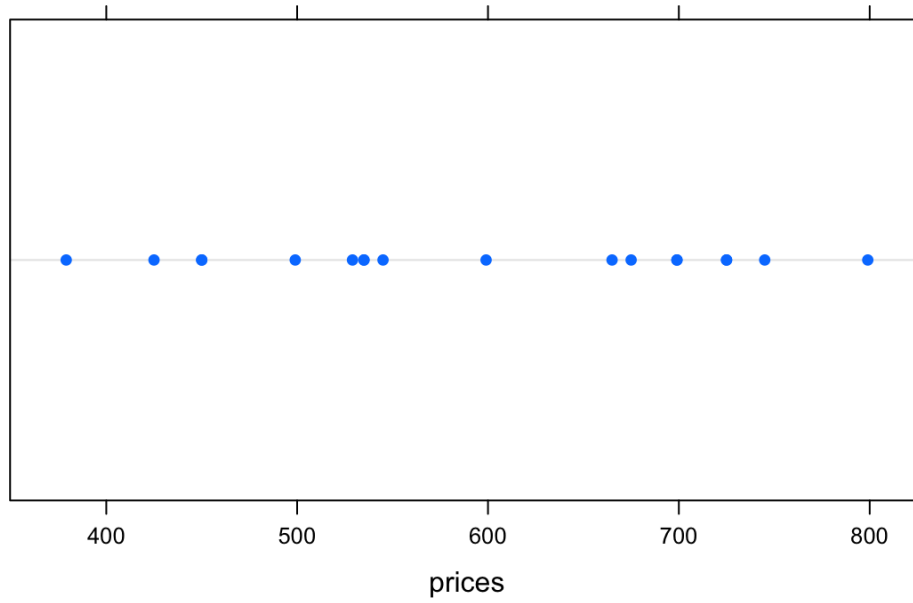
```
##
##   The decimal point is at the |
##
##  1 | 3333333333444444555555555555556666666677777788888888889999999999
##  2 | 0000000011112222233333334444445555555666666679999
##  3 | 01112233333344567889
##  4 | 0011112245555667779999
##  5 | 0001222333355788
##  6 | 012344
##  7 | 6
```


Stem and leaf plot, change the length (scale)

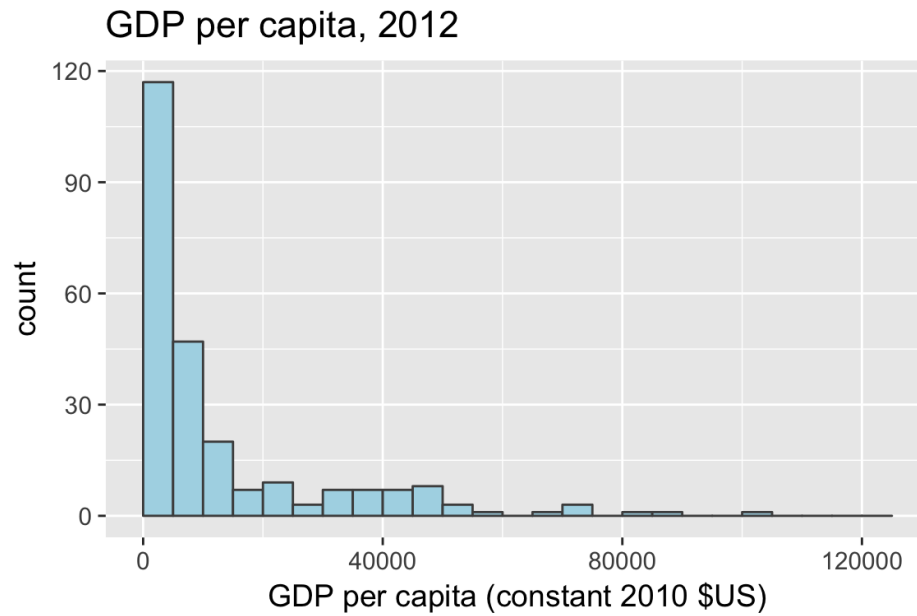
Total Fertility Rate

```
##
## The decimal point is at the |
##
## 1 | 333333333444444
## 1 | 5555555555555666666677777788888888889999999999
## 2 | 0000000011112222233333334444444
## 2 | 555555556666666679999
## 3 | 01112233333344
## 3 | 567889
## 4 | 001111224
## 4 | 5555667779999
## 5 | 00012223333
## 5 | 55788
## 6 | 012344
## 6 |
## 7 |
## 7 | 6
```

Dotplot

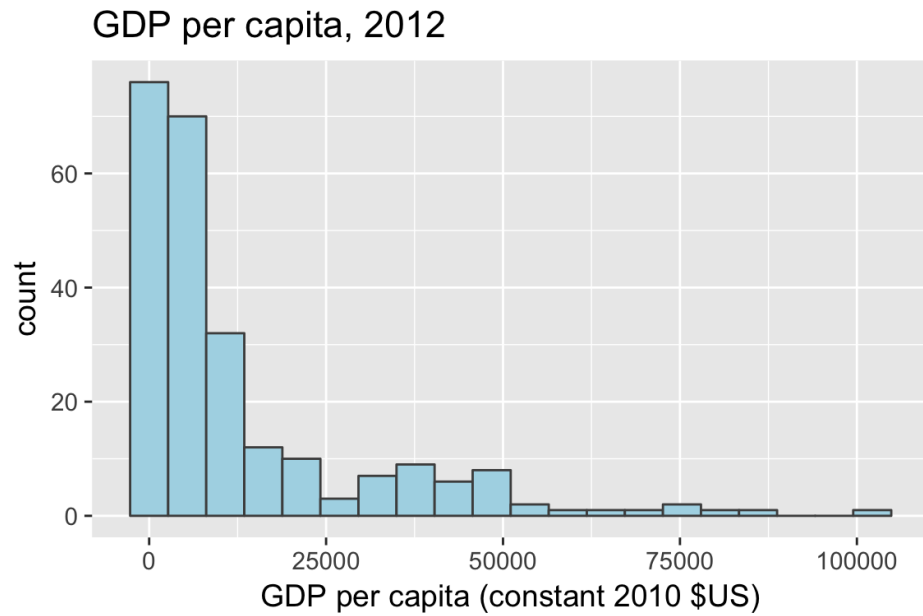


Frequency histogram

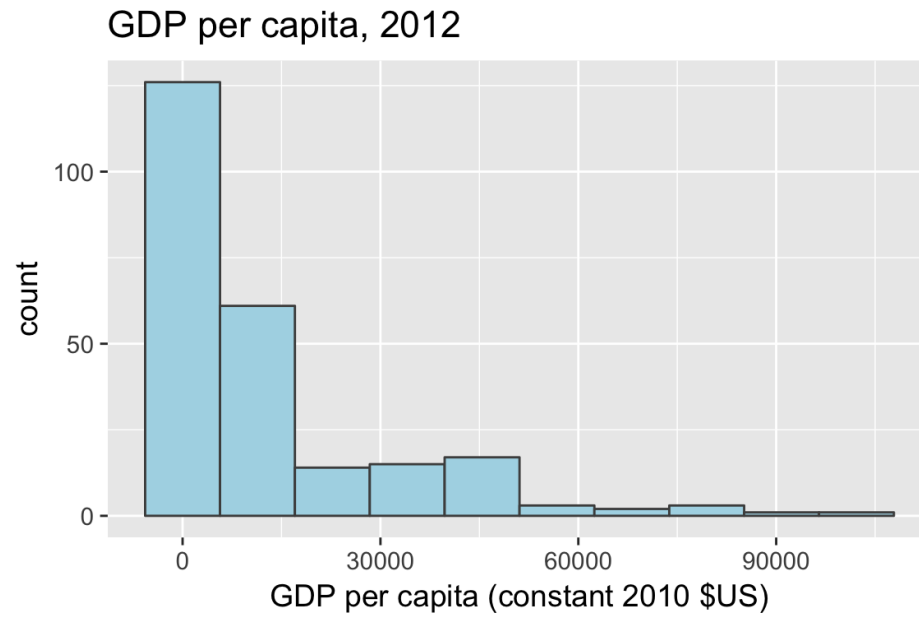


Source: World Bank

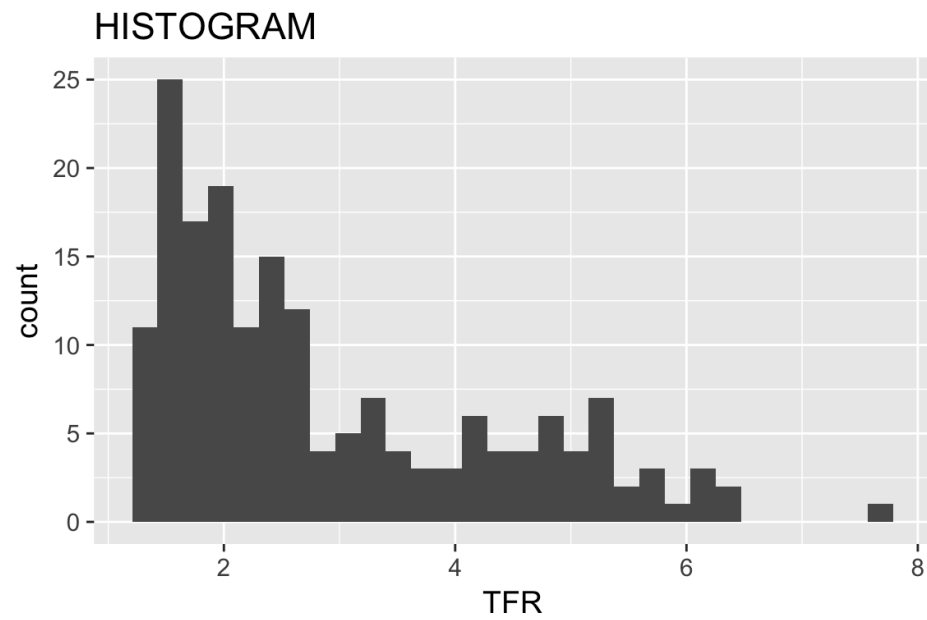
Change number of bins to 20



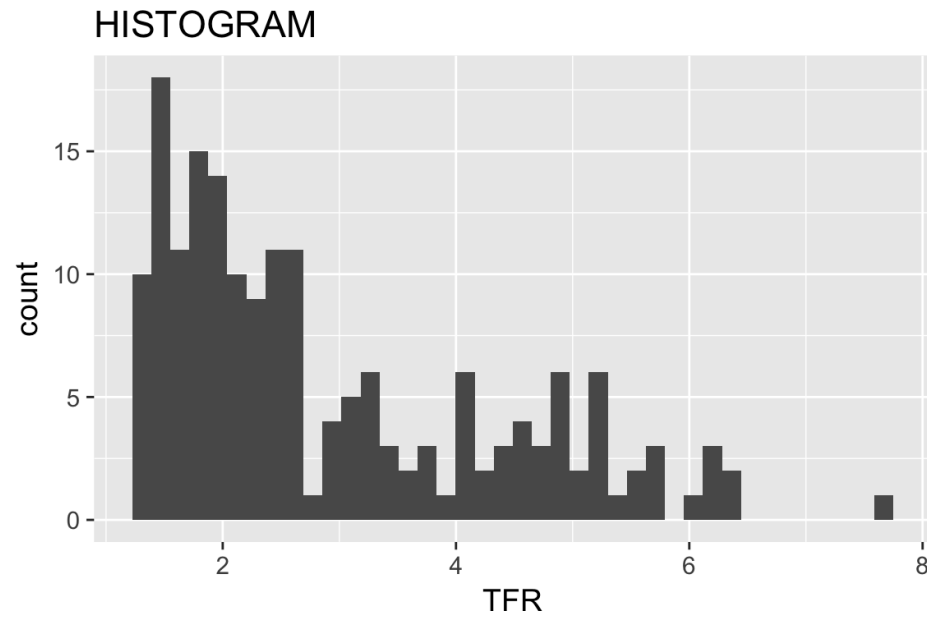
Change number of bins to 10



Total fertility rate



Change the number of bins to 40



Shapes of histograms

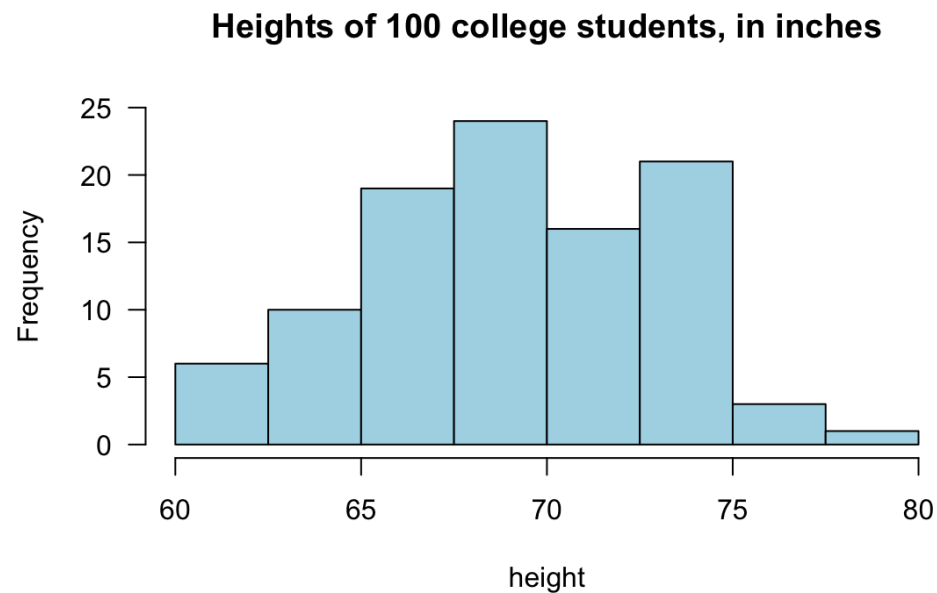
- Unimodal
- Bimodal
- Skewed left
- Skewed right

Discrete data

```
## [1] "Heights of 100 college students, in inches"
```

```
##      [1] 60 60 61 61 61 62 63 63 64 64 64 64 65 65 65
##     [16] 65 66 66 66 66 66 66 67 67 67 67 67 67 67 67
##     [31] 67 67 67 67 67 68 68 68 68 68 68 68 69 69 69
##     [46] 69 69 69 69 69 69 70 70 70 70 70 70 70 70 71
##     [61] 71 71 71 71 71 72 72 72 72 72 72 72 72 72 72
##     [76] 73 73 73 74 74 74 74 74 74 74 74 74 74 74 74
##     [91] 74 75 75 75 75 75 76 76 77 79
```

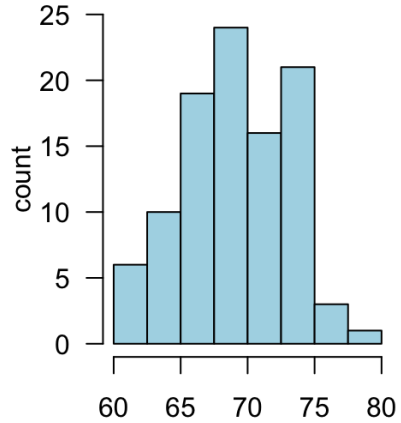
Discrete data



```
##      [1] 60 60 61 61 61 62 63 63 64 64 64 64 65 65 65
##     [16] 65 66 66 66 66 66 66 66 67 67 67 67 67 67 67
##     [31] 67 67 67 67 67 68 68 68 68 68 68 68 68 69 69 69
##     [46] 69 69 69 69 69 69 70 70 70 70 70 70 70 70 71
##     [61] 71 71 71 71 71 72 72 72 72 72 72 72 72 72 72
##     [76] 73 73 73 74 74 74 74 74 74 74 74 74 74 74 74
##     [91] 74 75 75 75 75 75 75 76 76 77 79
```

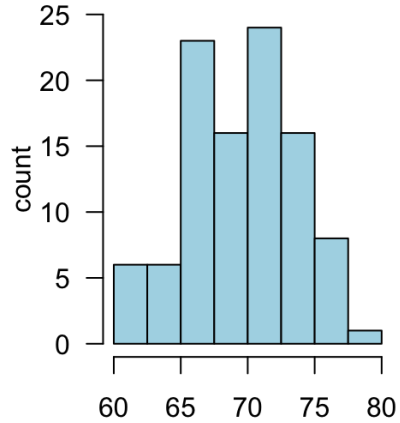
Discrete data histogram

Histogram of height



RIGHT CLOSED, LEFT OPEN

Histogram of height



RIGHT OPEN, LEFT CLOSED

##	[1]	60	60	61	61	61	62	63	63	64	64	64	64	65	65	65
##	[16]	65	66	66	66	66	66	66	67	67	67	67	67	67	67	67
##	[31]	67	67	67	67	67	68	68	68	68	68	68	68	69	69	69
##	[46]	69	69	69	69	69	69	70	70	70	70	70	70	70	70	71
##	[61]	71	71	71	71	71	72	72	72	72	72	72	72	72	72	72
##	[76]	73	73	73	74	74	74	74	74	74	74	74	74	74	74	74
##	[91]	74	75	75	75	75	75	76	76	77	79					

Discrete data – stem and leaf plot

```
##  
## The decimal point is at the |  
##  
## 60 | 00000  
## 62 | 000  
## 64 | 00000000  
## 66 | 00000000000000000000  
## 68 | 000000000000000000  
## 70 | 0000000000000000  
## 72 | 0000000000000000  
## 74 | 00000000000000000000  
## 76 | 000  
## 78 | 0
```

Discrete data – stem and leaf plot (new scale)

```
##  
##   The decimal point is 1 digit(s) to the right of the |  
##  
##   6 | 001112334444  
##   6 | 55556666667777777777777788888889999999999  
##   7 | 00000000111111222222222233344444444444444  
##   7 | 555556679
```