STAT UN1201 (002)

Prof. Joyce Robbins

January 17, 2017

Very brief course outline

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

Types of data

- Qualitative (categorical)
- Quantitative
  - *Discrete*
  - *Continuous*

Descriptive statistics

- numerical summaries: means, standard deviations, 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

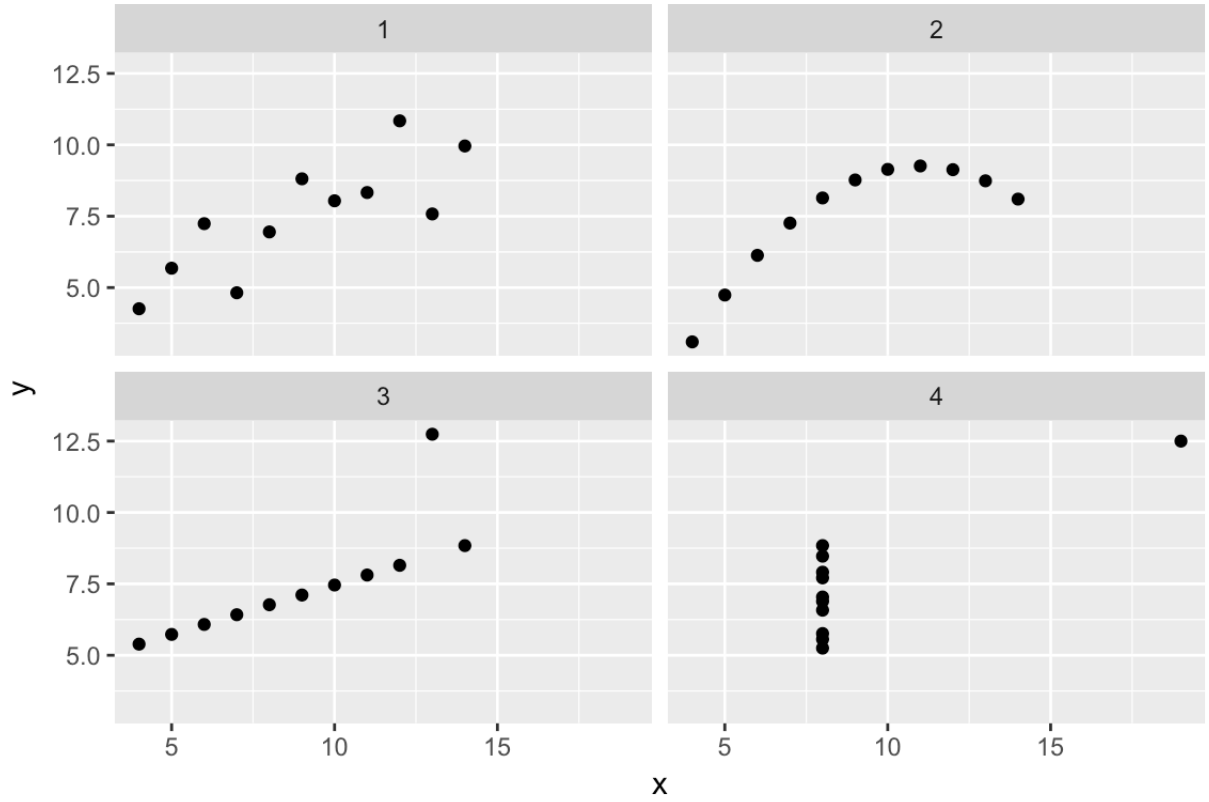
Anscombe's Quartet, numerical summaries

Each of the four data sets yields the same standard output from a typical regression program, namely

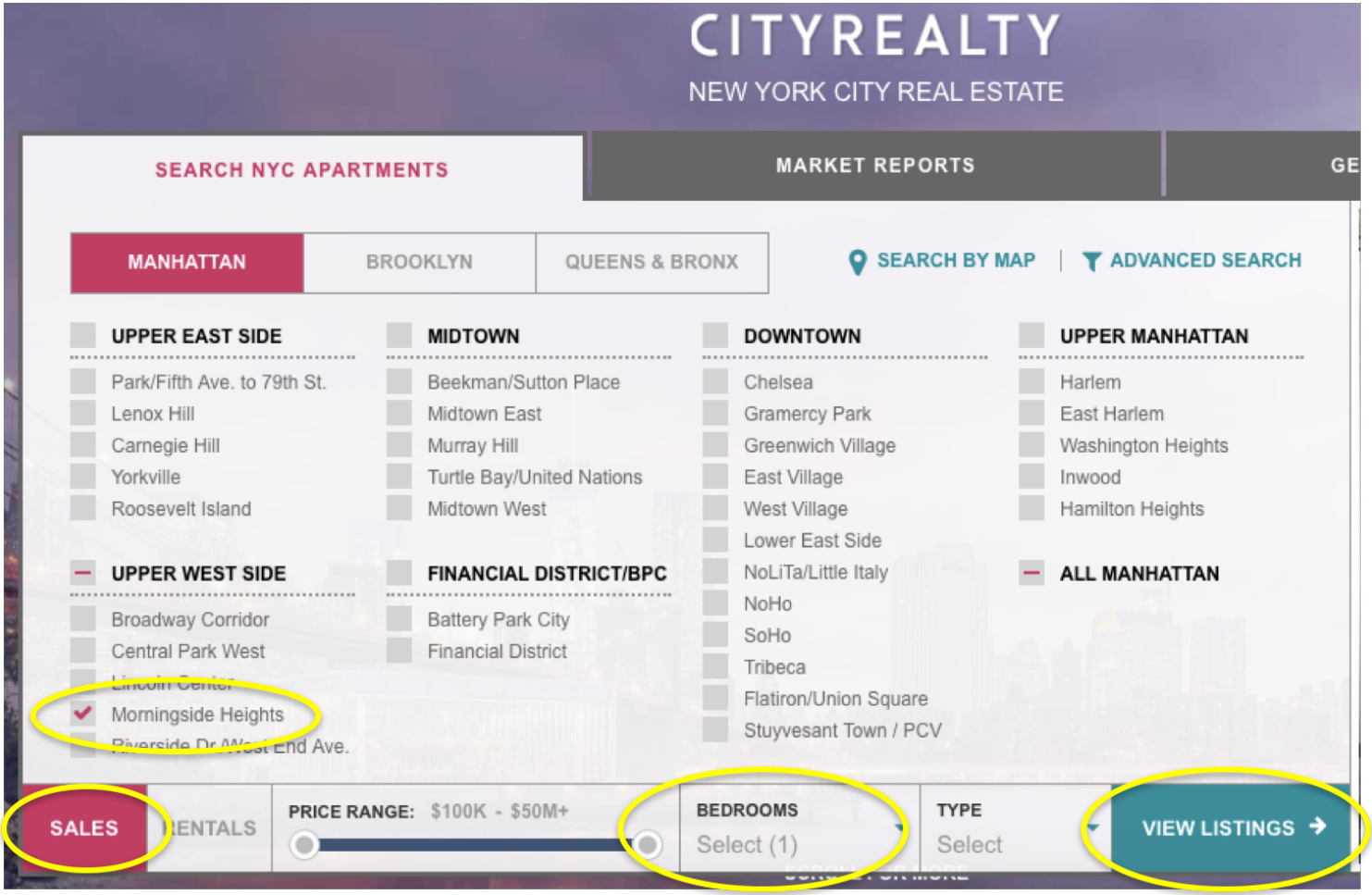
Number of observations $(n) = 11$
Mean of the x's $(\bar{x}) = 9.0$
Mean of the y's $(\bar{y}) = 7.5$
Regression coefficient (b_1) of y on $x = 0.5$
Equation of regression line: $y = 3 + 0.5\,x$
Sum of squares of $x - \bar{x} = 110.0$
Regression sum of squares $= 27.50$ (1 d.f.)
Residual sum of squares of $y = 13.75$ (9 d.f.)
Estimated standard error of $b_1 = 0.118$
Multiple $R^2 = 0.667$

F.J. Anscombe, "Graphs in Statistical Analysis"

Anscombe's Quartet, graphical analysis



Collecting data

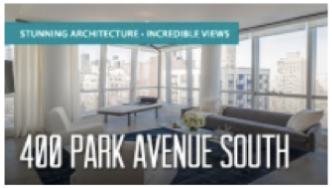


Collecting data

HIDE ALL PHOTOS SHOW ALL PHOTOS

ADDRESS	TYPE	NEIGHBORHOOD	BEDS / BATH / SIZE	PRICE	PRICE/FT²	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 ft²	$745K	$931	Jul 19, 2016
380 Riverside Drive, #8L	CO-OP	Morningside Heights	1 bd / 1 ba / 700 ft²	$725K	$1,036	Sep 06, 2016
535 West 110th Street, #6B	CO-OP	Morningside Heights	1 bd / 1 ba / 750 ft²	$725K	$967	Sep 07, 2016
535 West 110th Street, #6A	CO-OP	Morningside Heights	1 bd / 1 ba / 750 ft²	$699K (-4%)	$933	Apr 26, 2016
395 Riverside Drive, #2D	CO-OP	Morningside Heights	1 bd / 1 ba / 900 ft²	$699K	$777	Mar 03, 2016
528 West 111th Street, #24	CO-OP	Morningside Heights	1 bd / 1 ba	$675K (-6.9%)	-	Apr 15, 2015
545 West 111th Street, #9N	CO-OP	Morningside Heights	1 bd / 1 ba / 650 ft²	$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 ft²	$535K	$892	Jun 22, 2016
80 La Salle Street, #21B	CO-OP	Morningside Heights	1 bd / 1 ba / 725 ft²	$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 ft²	$450K	$900	Jun 28, 2016

TURED NEW DEVELOPMENTS



510 West 123rd Street, #38	CO-OP	Morningside Heights	1 bd / 1 ba / 700 ft²	$450K	$643	Jul 02, 2016
3115 Broadway, #43	CO-OP	Morningside Heights	1 bd / 1 ba / 650 ft²	$425K	$654	Apr 18, 2016
3115 Broadway, #42	CO-OP	Morningside Heights	1 bd / 1 ba / 650 ft²	$379K (-5%)	$583	Mar 28, 2016

Cleaning / organizing data

	A	B	C	D	E	F	G	H
1		ADDRESS	TYPE	NEIGHBORH	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				

Drawing stem-and-leaf plots

Prices:

```
##  [1] 379 425 450 450 499 529 535 535 545 599 665 675 699 699 725 725 745
## [18] 799
```

Prices, rounded to 2 significant digits:

```
##  [1] 380 420 450 450 500 530 540 540 540 600 660 680 700 700 720 720 740
## [18] 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)

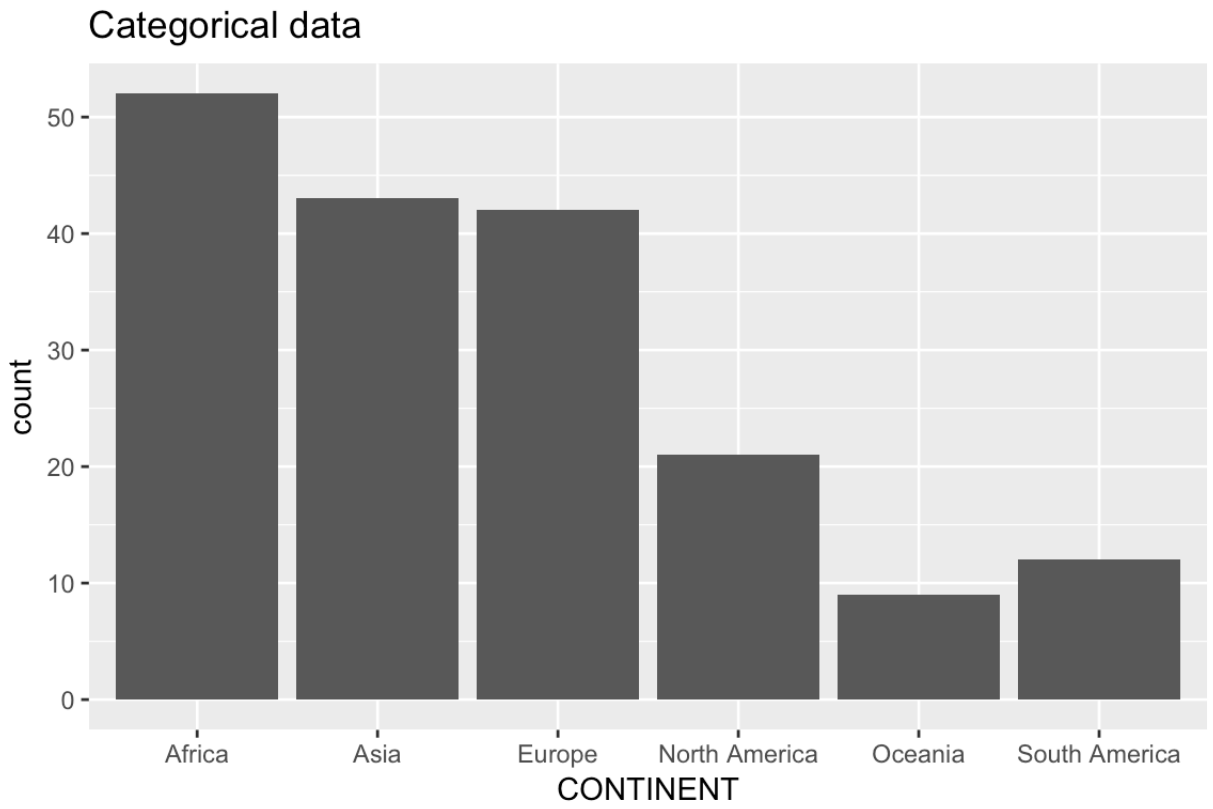
```
##
##   The decimal point is at the |
##
##   1 | 333333333444444555555555555566666666777777888888888899999999999
##   2 | 000000001111222223333333444444455555555666666679999
##   3 | 01112233333344567889
##   4 | 0011112245555667779999
##   5 | 0001222333355788
##   6 | 012344
##   7 | 6
```

Stem and leaf plot, change the length (scale)

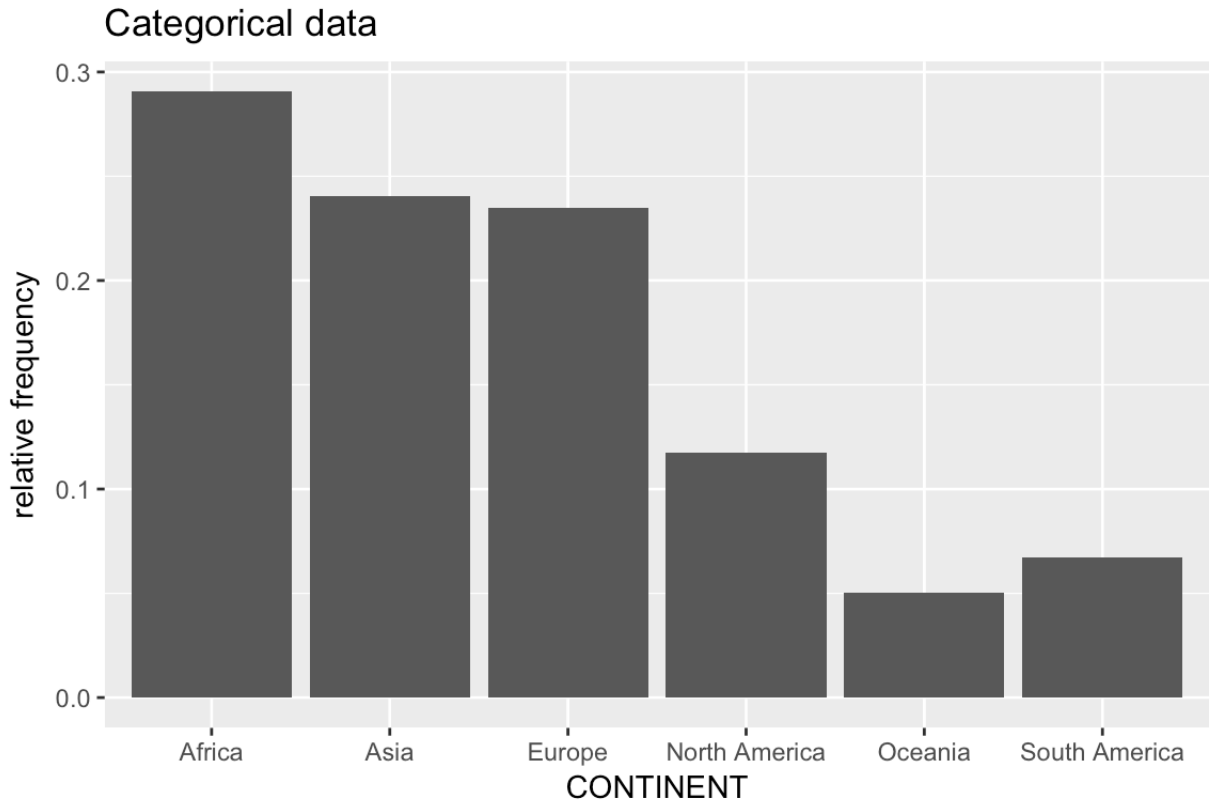
Total Fertility Rate

```
##
##   The decimal point is at the |
##
##    1 | 333333333444444
##    1 | 55555555555556666666677777788888888889999999999
##    2 | 00000000111122222333333444444444
##    2 | 555555556666666679999
##    3 | 01112233333344
##    3 | 567889
##    4 | 001111224
##    4 | 5555667779999
##    5 | 00012223333
##    5 | 55788
##    6 | 012344
##    6 |
##    7 |
##    7 | 6
```

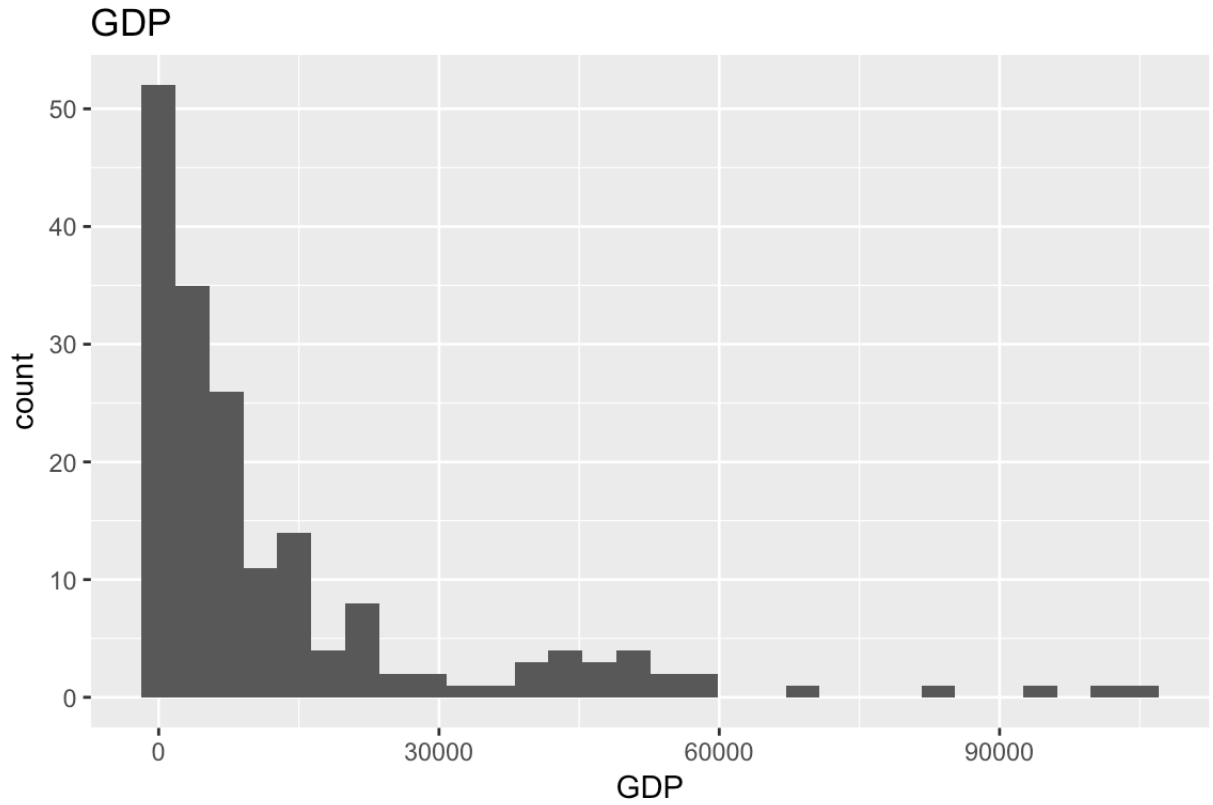
Bar chart


Categorical data
count
50
40
30
20
10
0
Africa
Asia
Europe
North America
Oceania
South America
CONTINENT

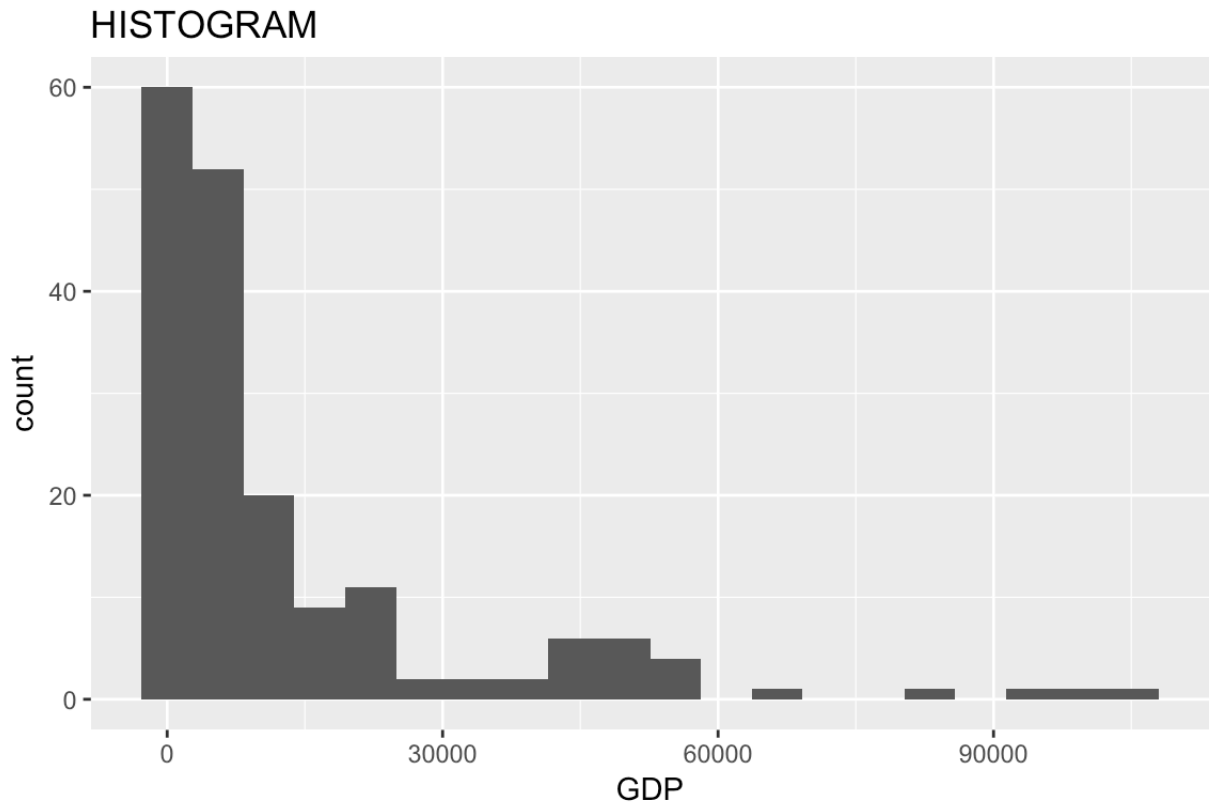
Relative frequency



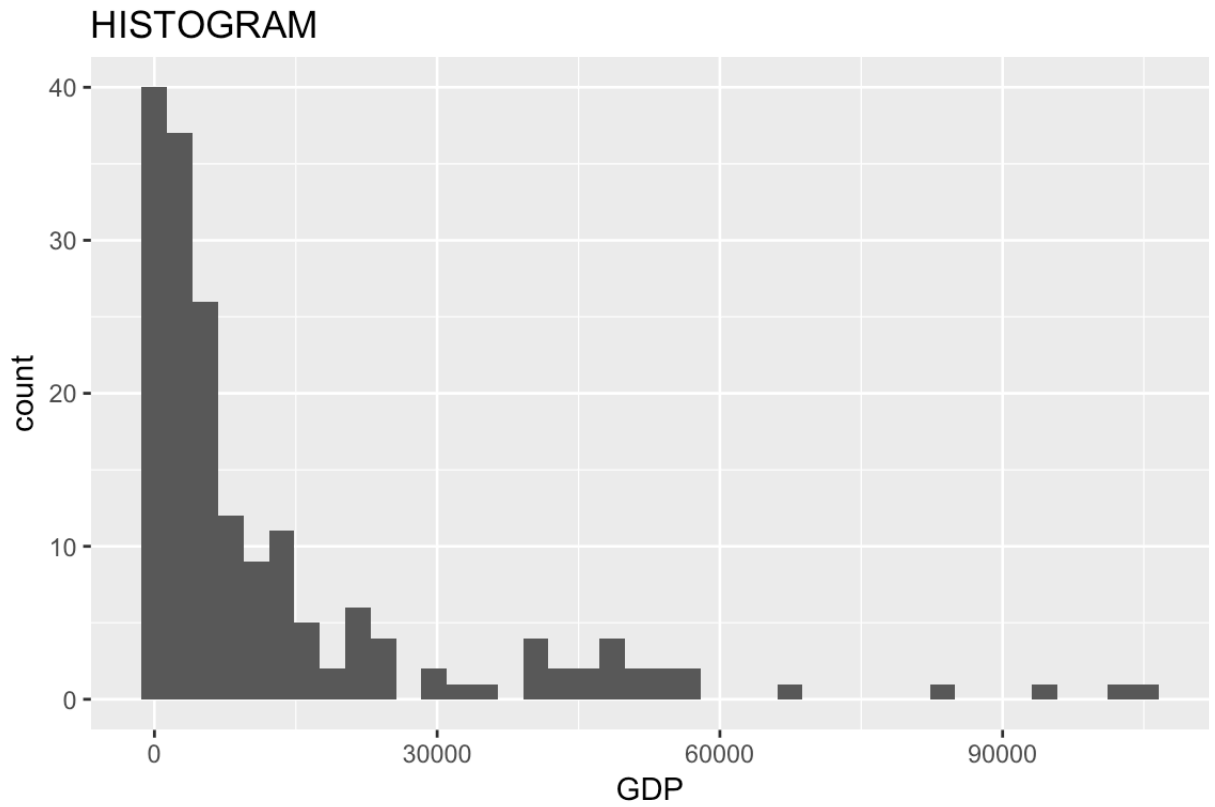
Histogram



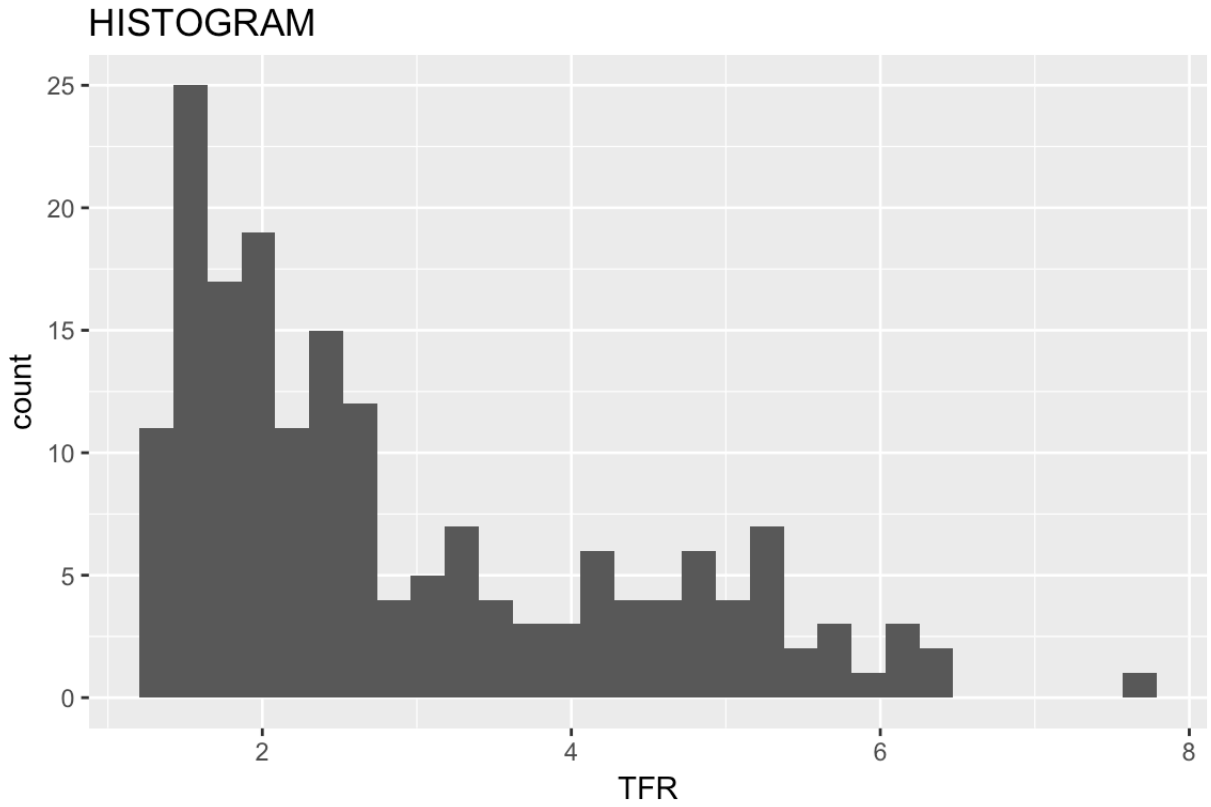
Change number of bins to 20



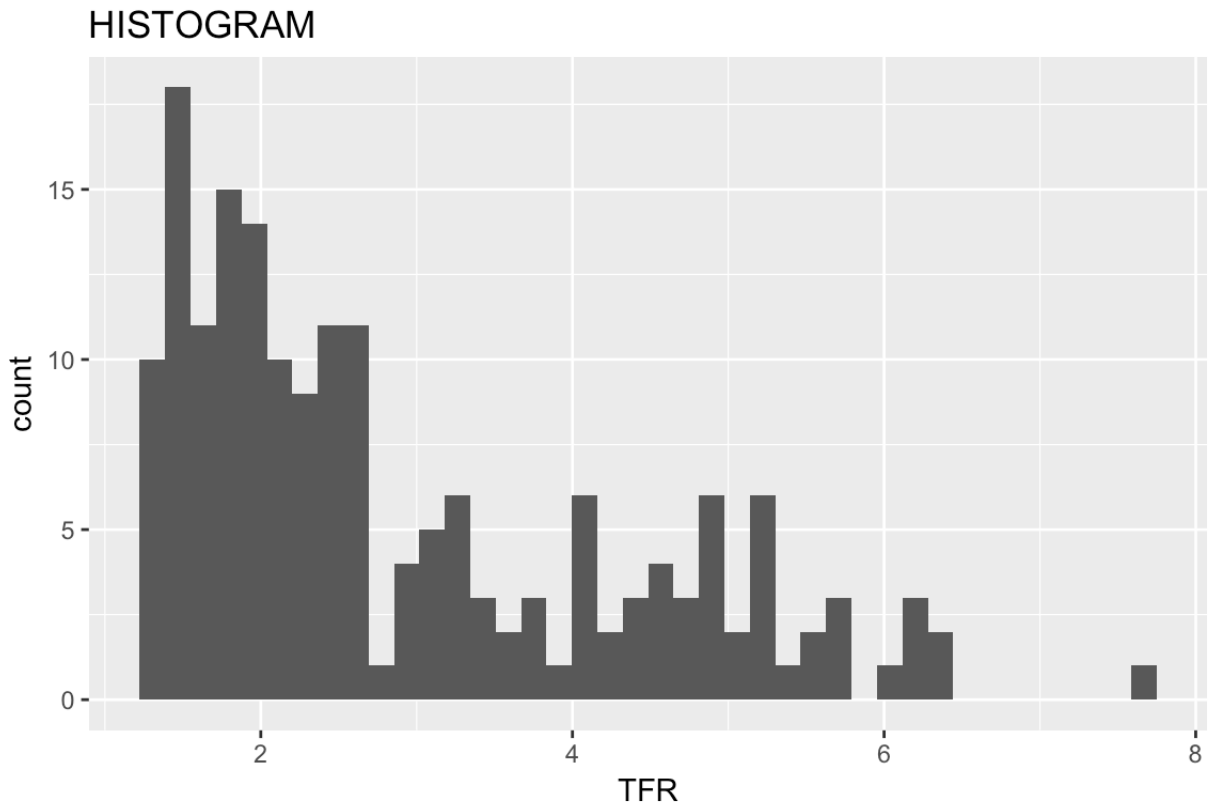
Change number of bins to 40



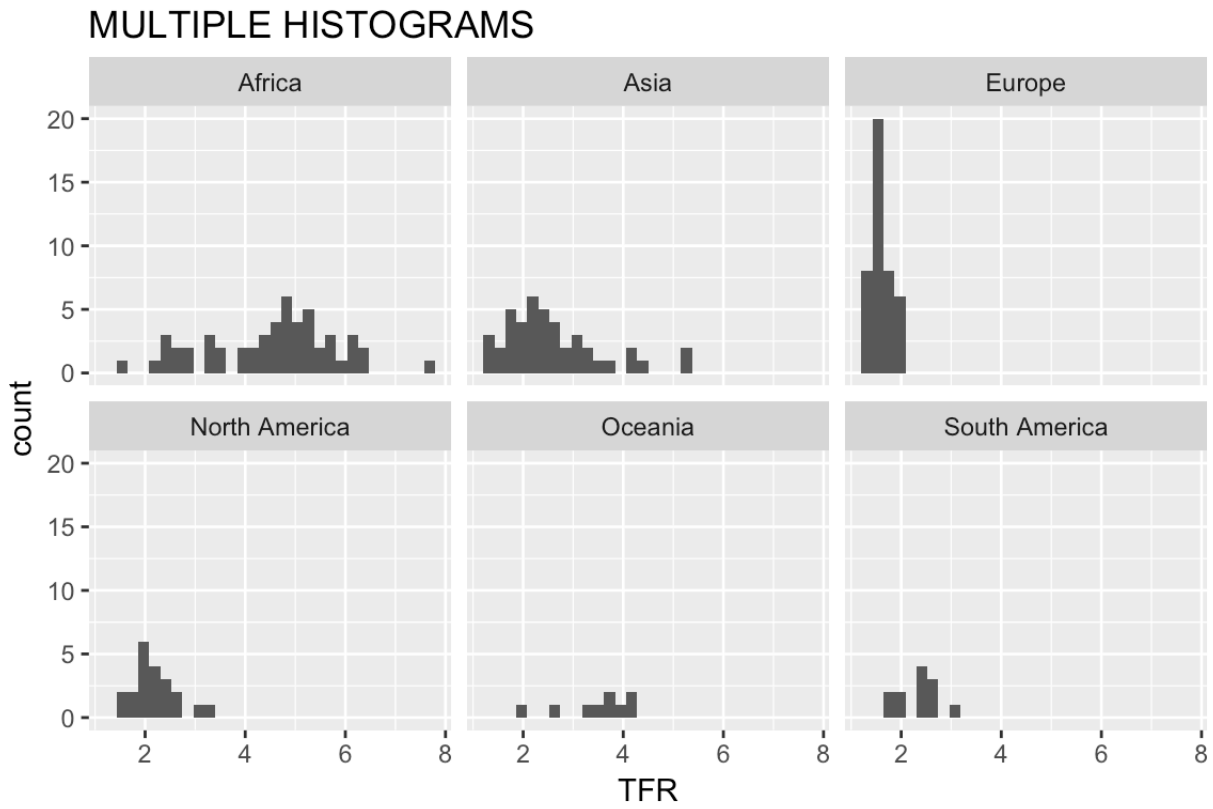
Another continuous variable: total fertility rate



Change the number of bins to 40



Break down the results by continent



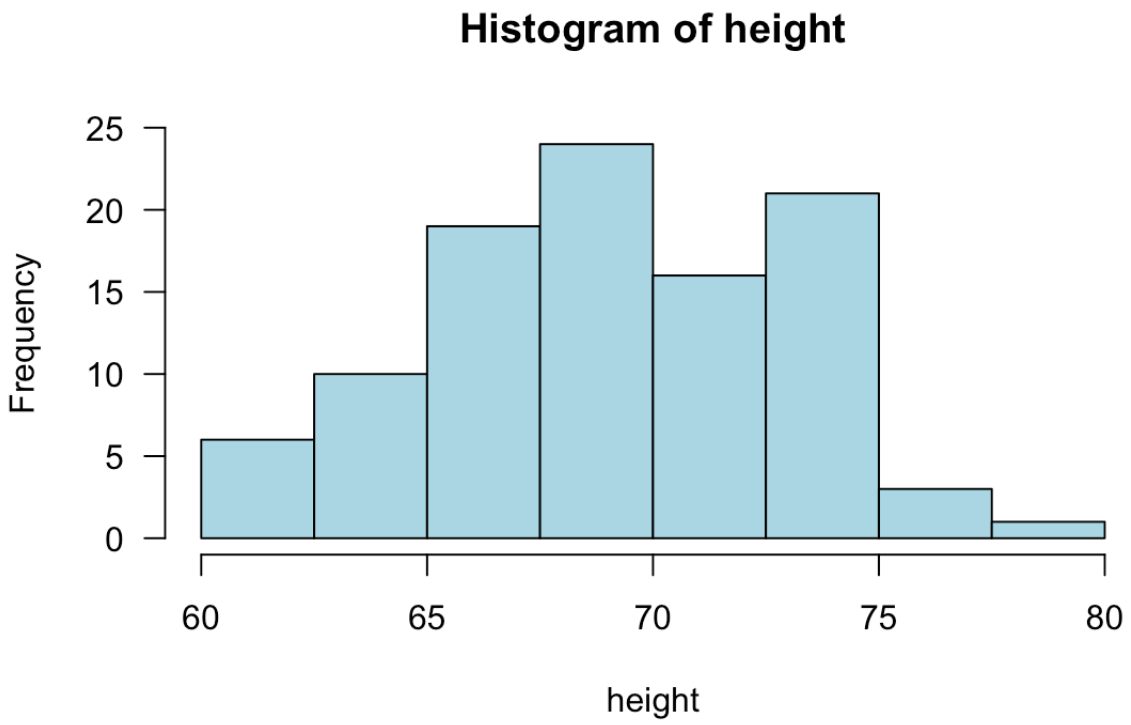
Shapes of histograms

- Unimodal
- Bimodal
- Skewed left
- Skewed right

Quantitative data – discrete

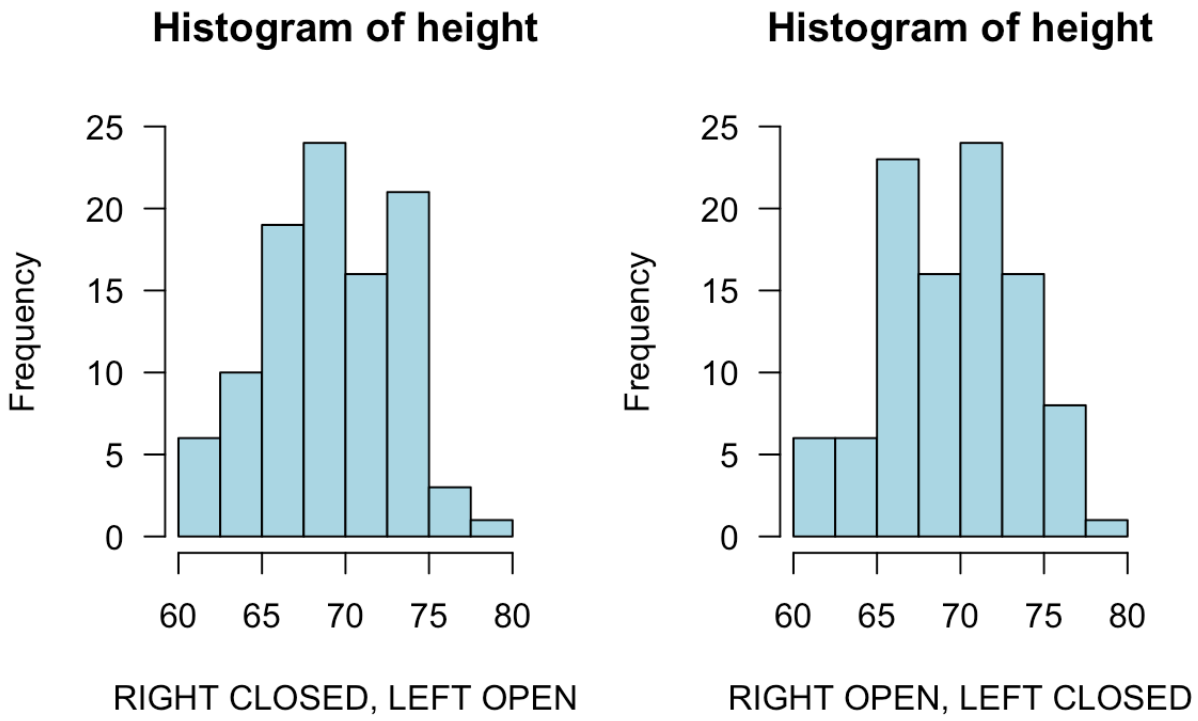
```
## [1] "Heights of 100 college students, in inches"
```

```
##   [1] 60 60 61 61 61 62 63 63 64 64 64 64 65 65 65 65 66 66 66 66 66 66 67
##  [24] 67 67 67 67 67 67 67 67 67 67 67 68 68 68 68 68 68 68 69 69 69 69
##  [47] 69 69 69 69 70 70 70 70 70 70 70 70 71 71 71 71 71 71 72 72 72 72
##  [70] 72 72 72 72 72 73 73 73 74 74 74 74 74 74 74 74 74 74 74 74 74 75
##  [93] 75 75 75 75 76 76 77 79
```



Discrete data histograms

```
##   [1] 60 60 61 61 61 62 63 63 64 64 64 64 65 65 65 65 66 66 66 66 66 66 67
##  [24] 67 67 67 67 67 67 67 67 67 67 67 68 68 68 68 68 68 68 69 69 69 69
##  [47] 69 69 69 69 70 70 70 70 70 70 70 70 71 71 71 71 71 71 72 72 72 72
##  [70] 72 72 72 72 72 73 73 73 74 74 74 74 74 74 74 74 74 74 74 74 74 75
##  [93] 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 | 0000000000000000
##   70 | 00000000000000
##   72 | 0000000000000
##   74 | 000000000000000000
##   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 | 55556666667777777777778888888999999999
##   7 | 00000000111111222222222333444444444444444
##   7 | 555556679
```

Drawing histograms

Asking prices for one-bedroom apartments in Morningside Heights, in thousands
Data source: cityrealty.com, 9/13/2016

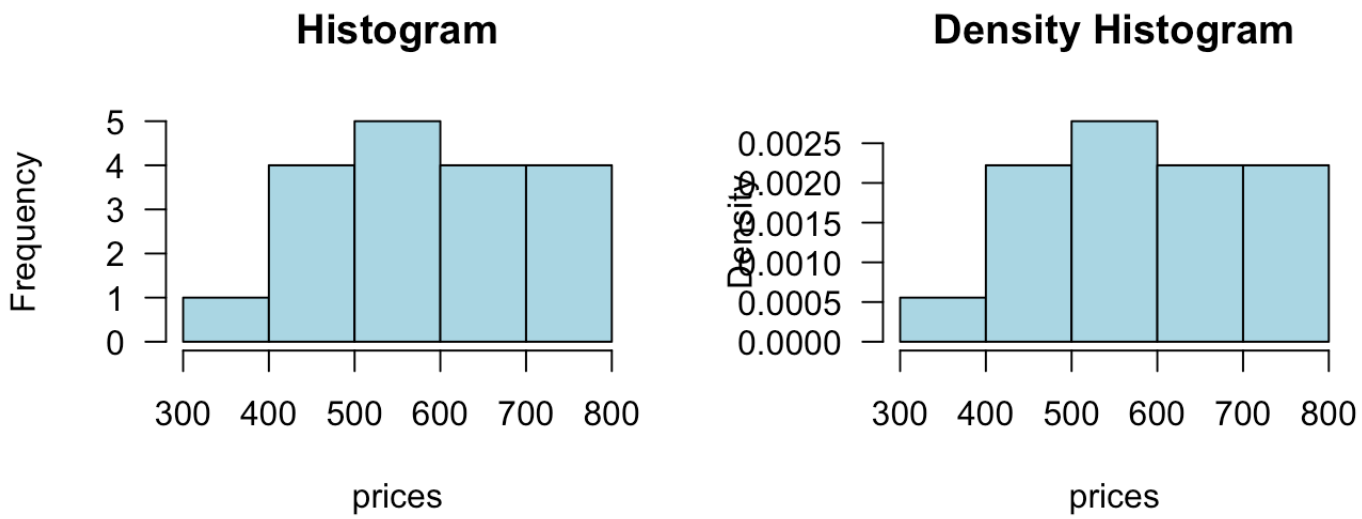
379, 425, 450, 450, 499, 529, 535, 535, 545,
599, 665, 675, 699, 699, 725, 725, 745, 799

Drawing histograms

Asking prices for one-bedroom apartments in Morningside Heights, in thousands
Data source: cityrealty.com, 9/13/2016

379, 425, 450, 450, 499, 529, 535, 535, 545,
599, 665, 675, 699, 699, 725, 725, 745, 799

Density histogram



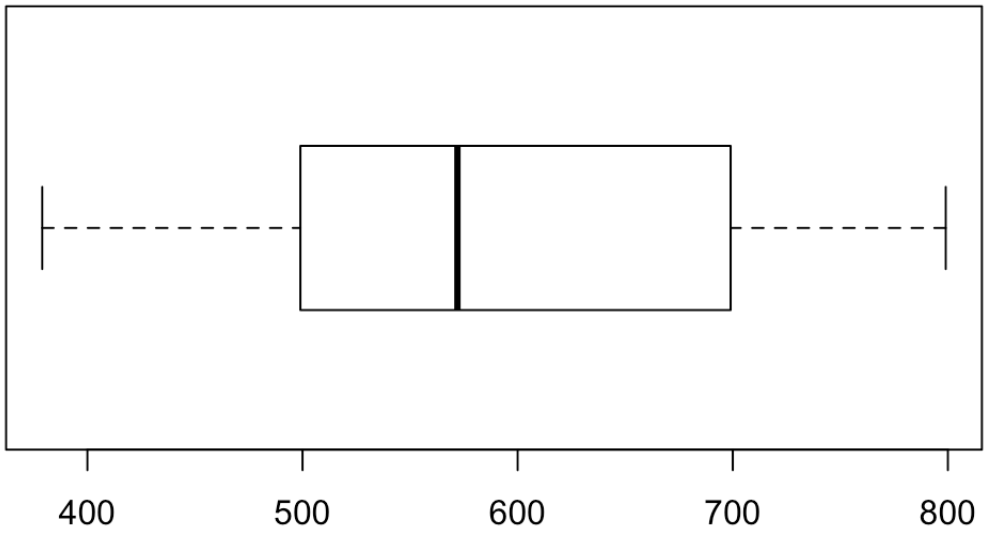
Class	Frequency	Rel. Frequency	Density
(300, 400]	1	.056	.00056
(400, 500]	4	.222	.00222
(500, 600]	5	.278	.00278
(600, 700]	4	.222	.00222
(700, 800]	4	.222	.00222

Five number summary

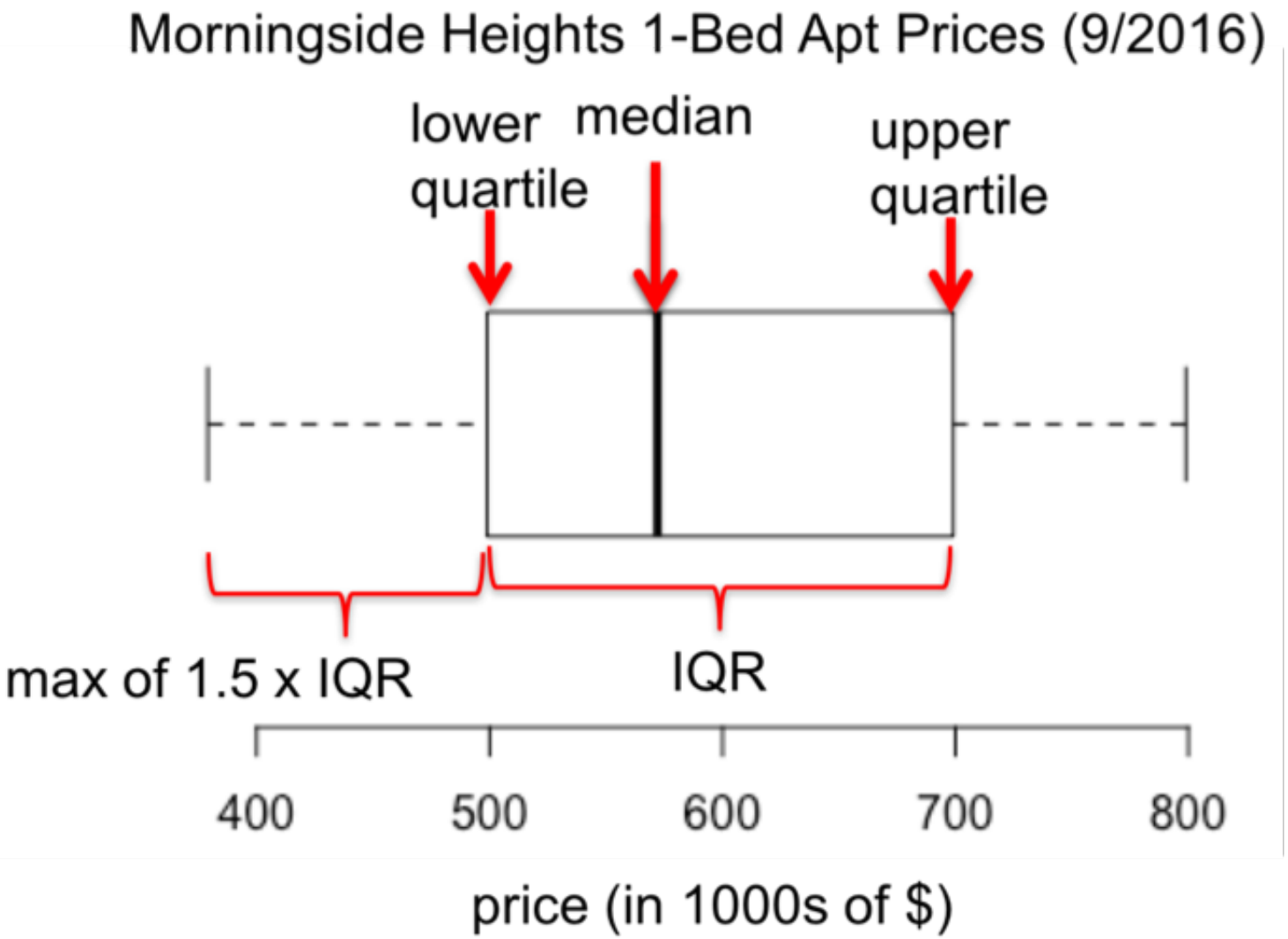
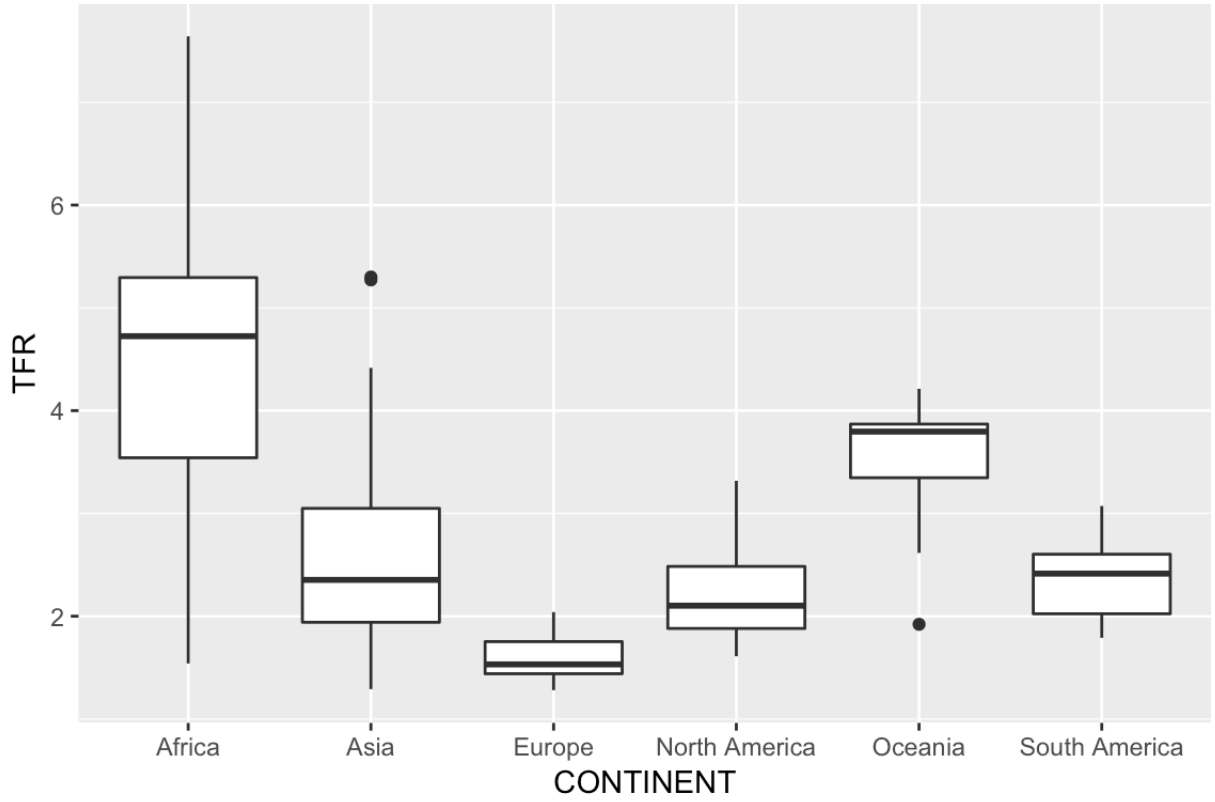
1. minimum
2. maximum
3. middle number (median)
4. 1st quantile
5. 3rd quantile

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   379.0   506.5   572.0   593.2   699.0   799.0
```

Box plot



Box plot


Morningside Heights 1-Bed Apt Prices (9/2016)
lower quartile
median
upper quartile
max of 1.5 x IQR
IQR
400
500
600
700
800
price (in 1000s of $)

Multiple box plots