

Continuous Variables

Continuous Variables

We're looking for features such as:

- Asymmetry
- Outliers
- Multimodality
- Gaps
- Heaping / Rounding
- Impossibilities / Errors

Histograms

- primary tool for continuous data
- count / relative frequency / density
- boundaries
- binwidth

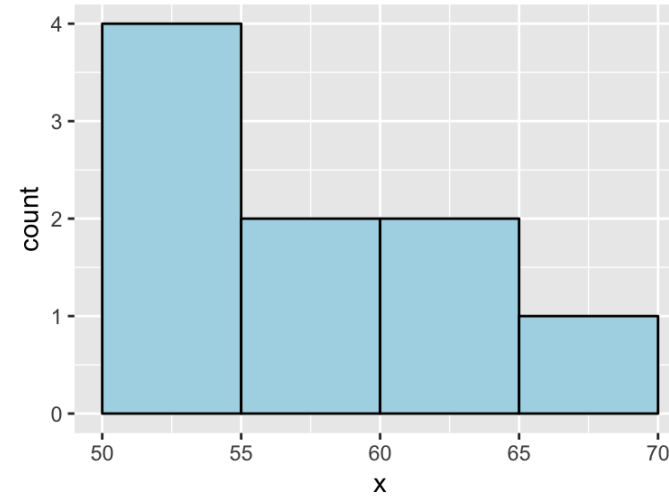
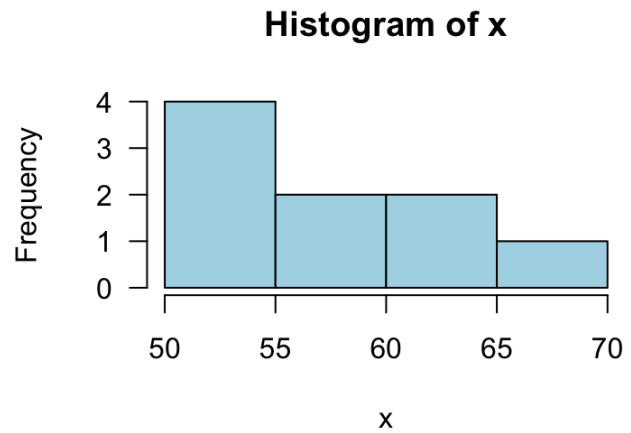
How are histograms created?

Draw a histogram on paper of the following data.

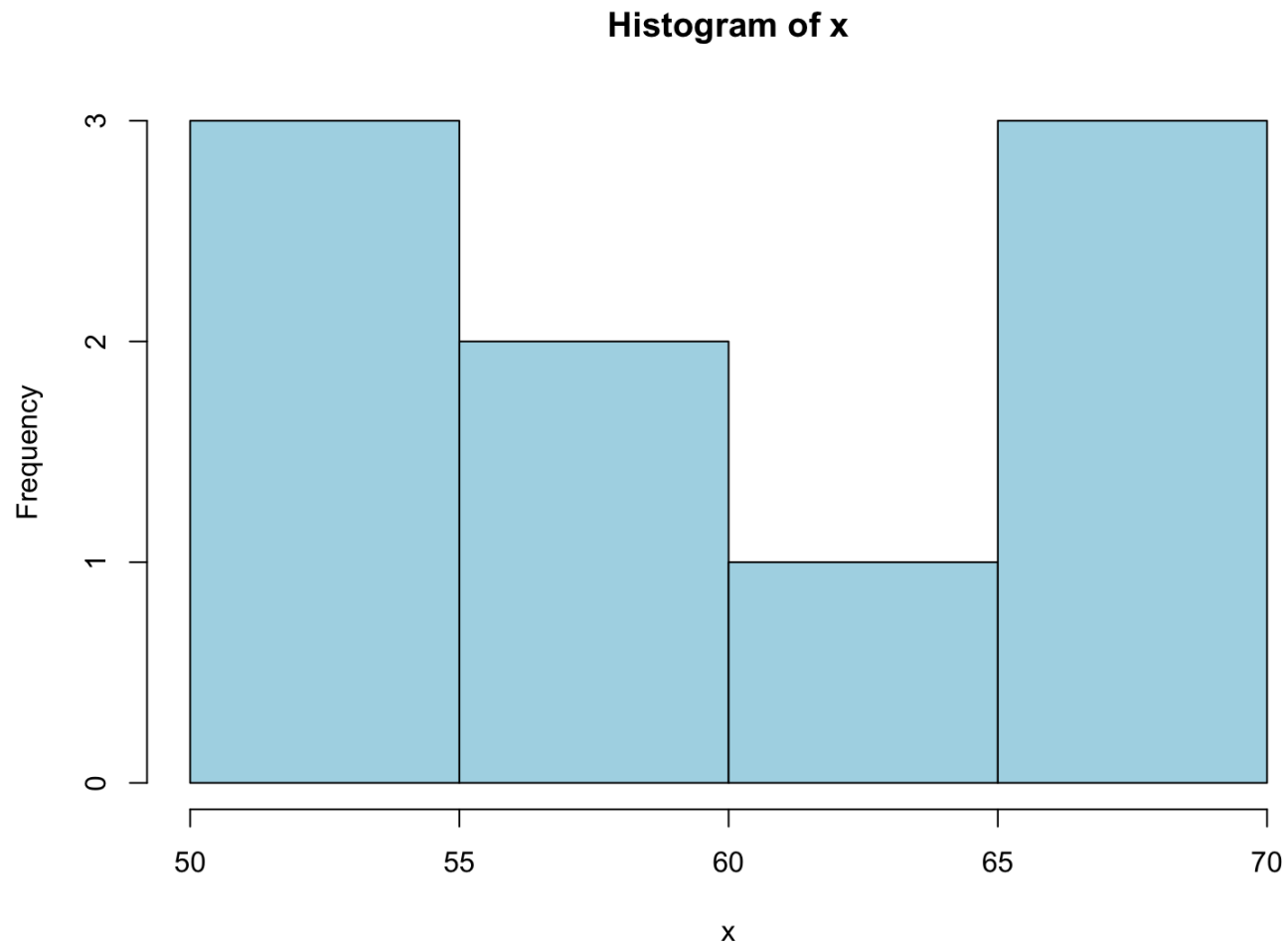
(use binwidth = 5)

50, 51, 53, 55, 56, 60, 65, 65, 68

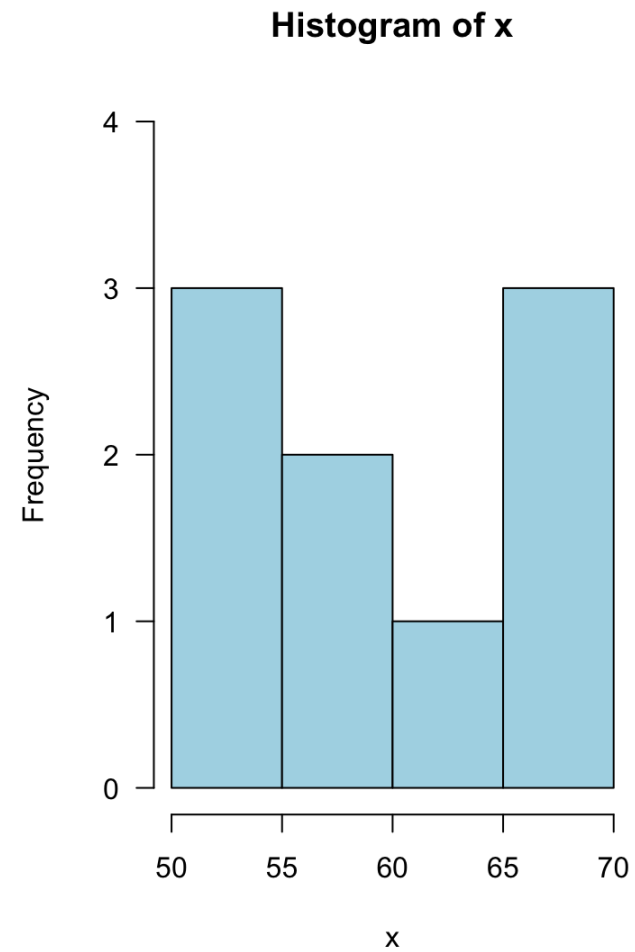
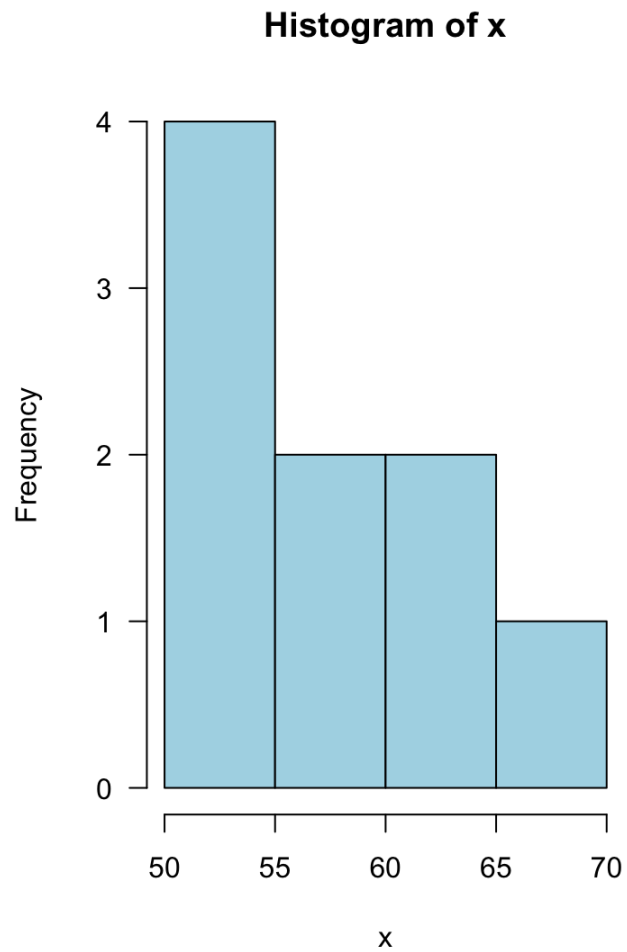
How are histograms created?



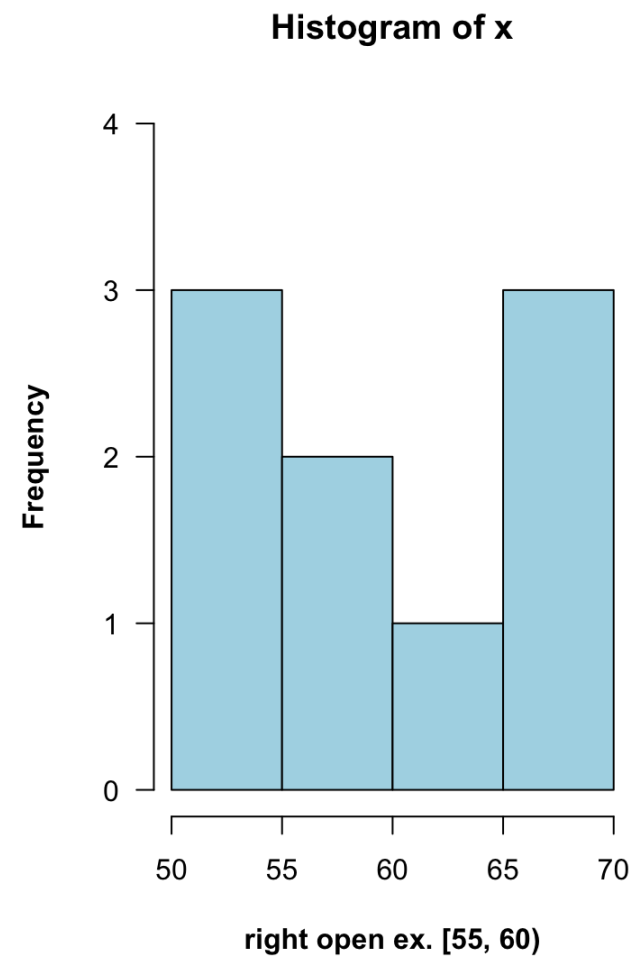
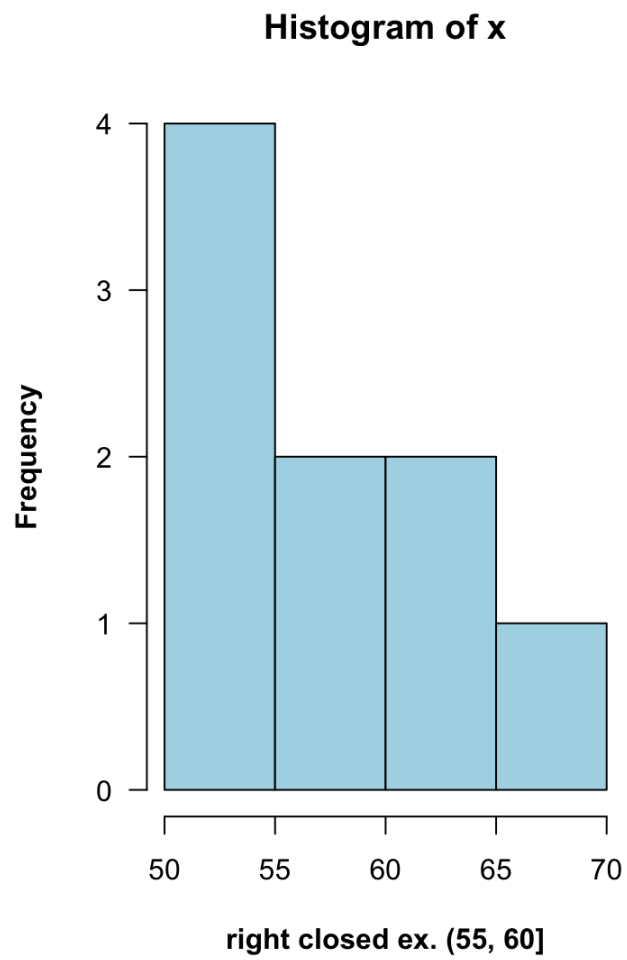
How are histograms created?



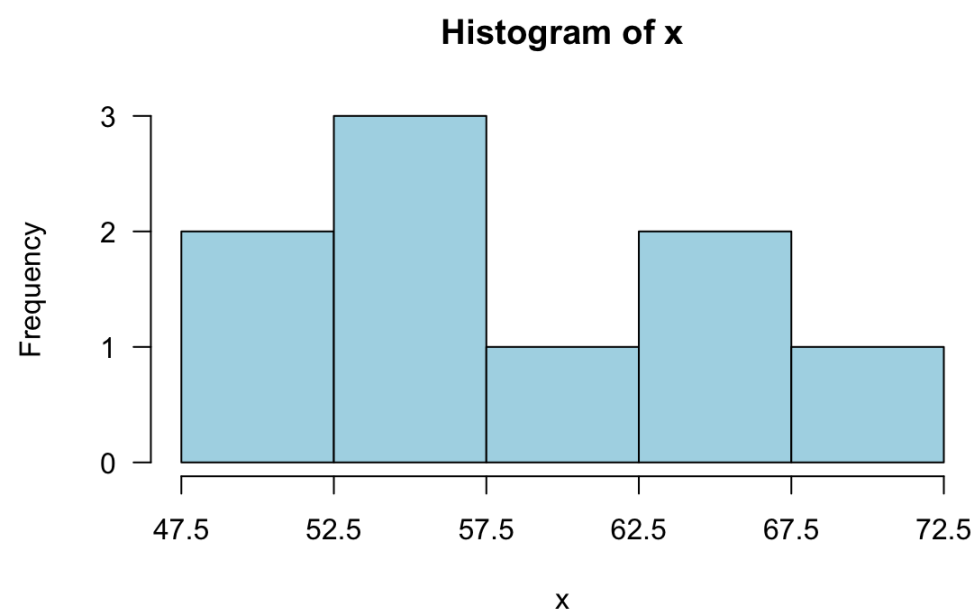
What is causing the difference?



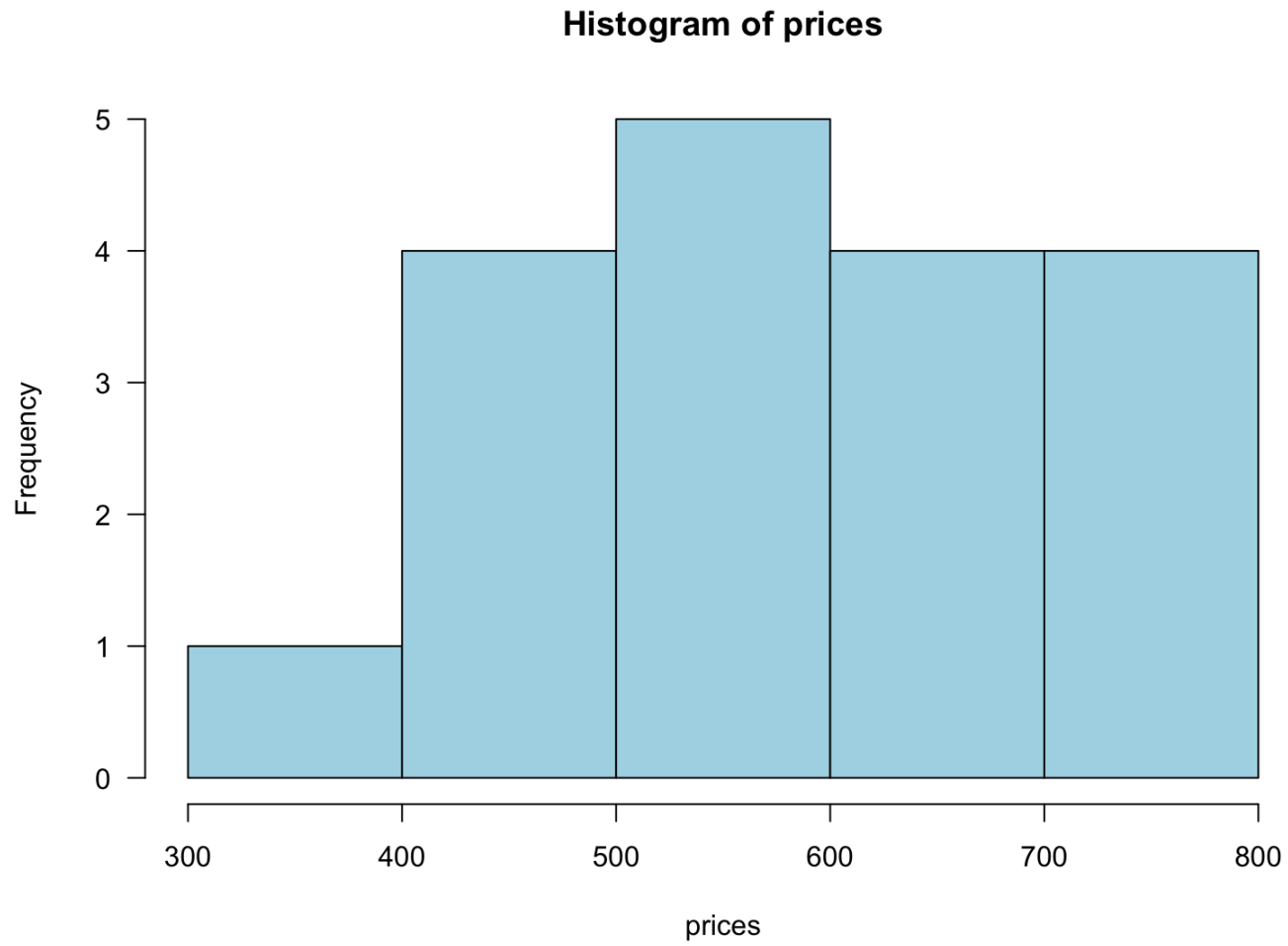
Bin boundaries



Bin boundaries

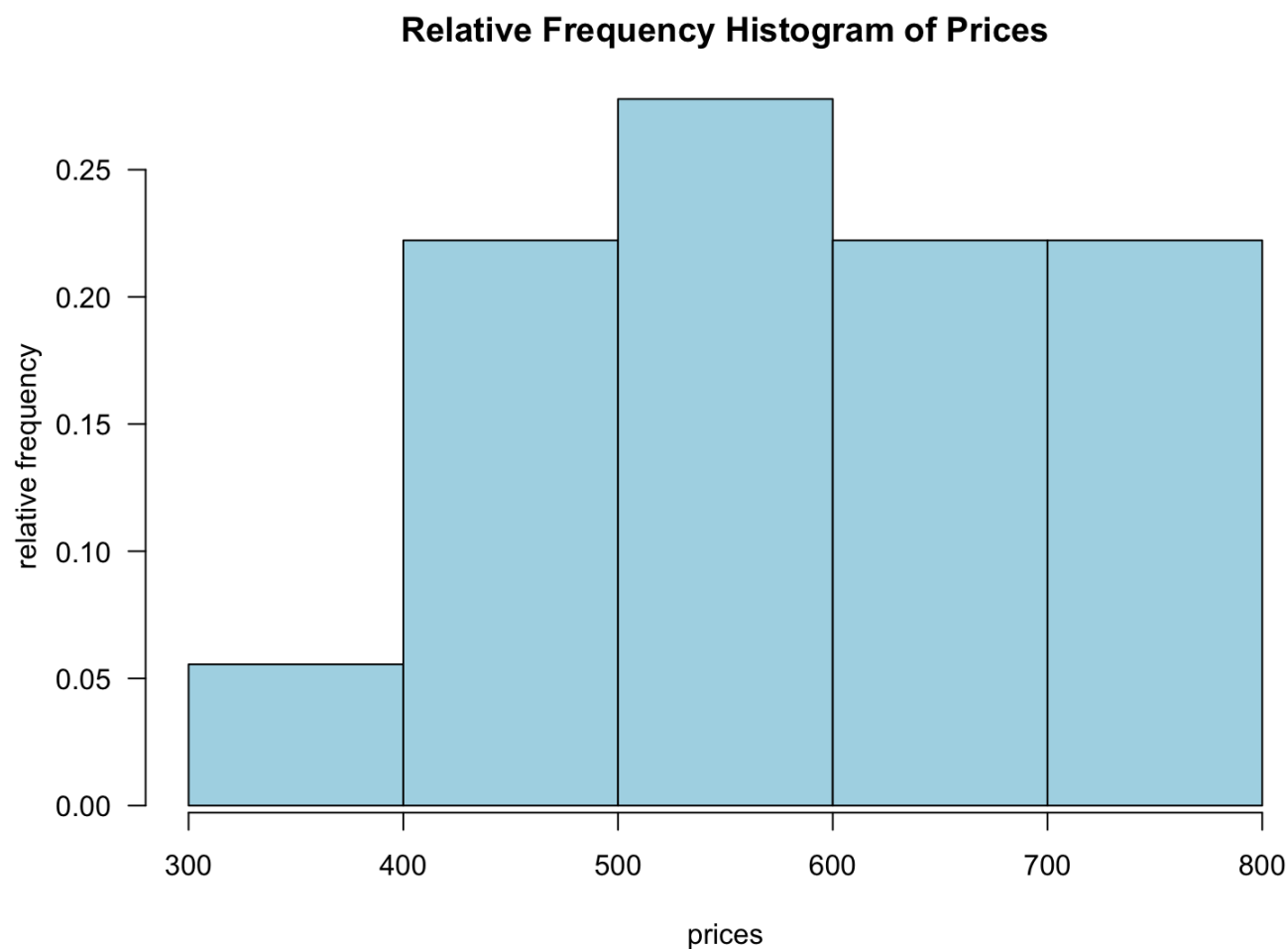


Frequency (count) histogram

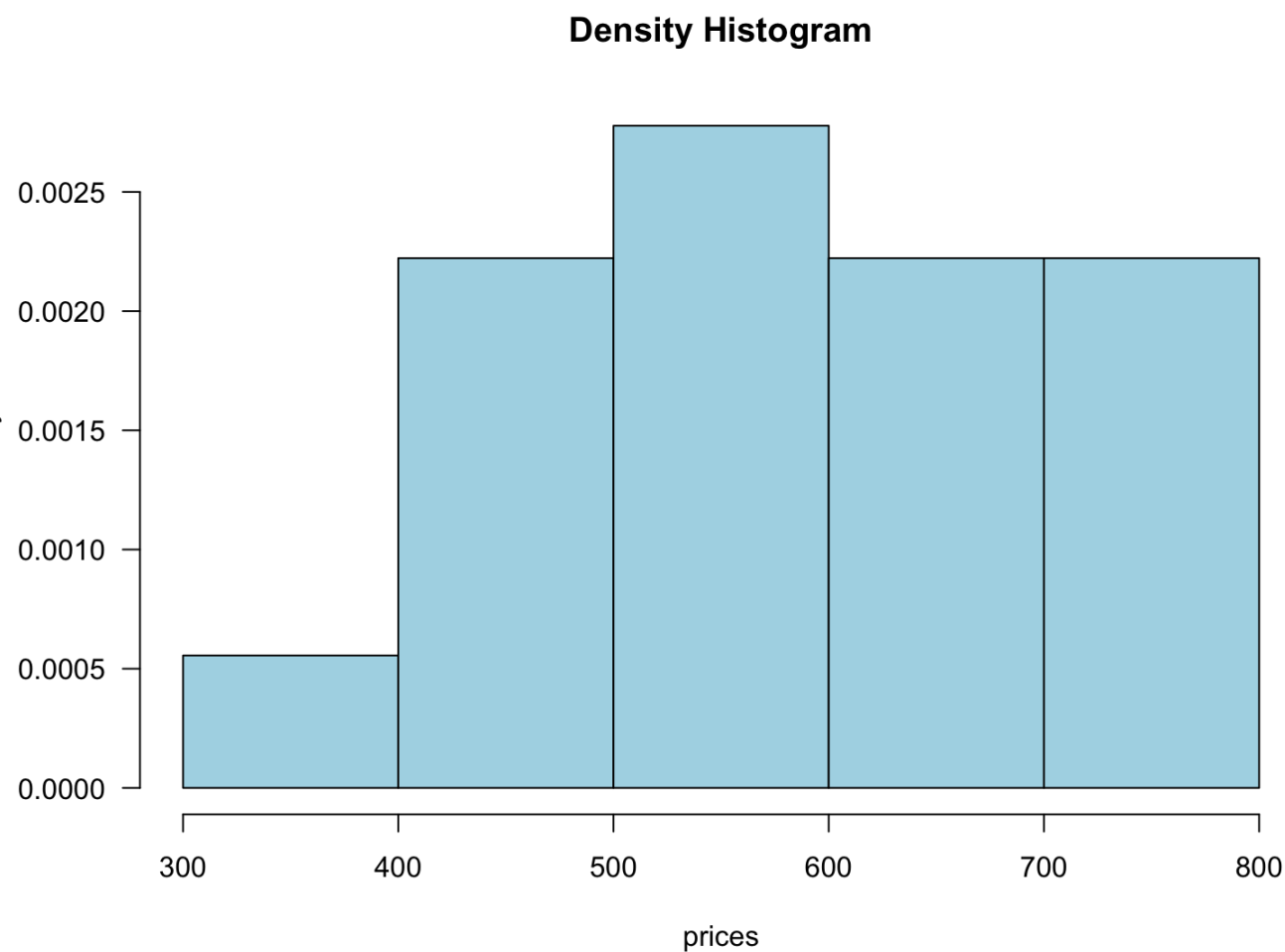


Prices of one-bedroom apartments in Morningside Heights (zip 10027) in \$1000k

Relative frequency histogram



Density Histogram

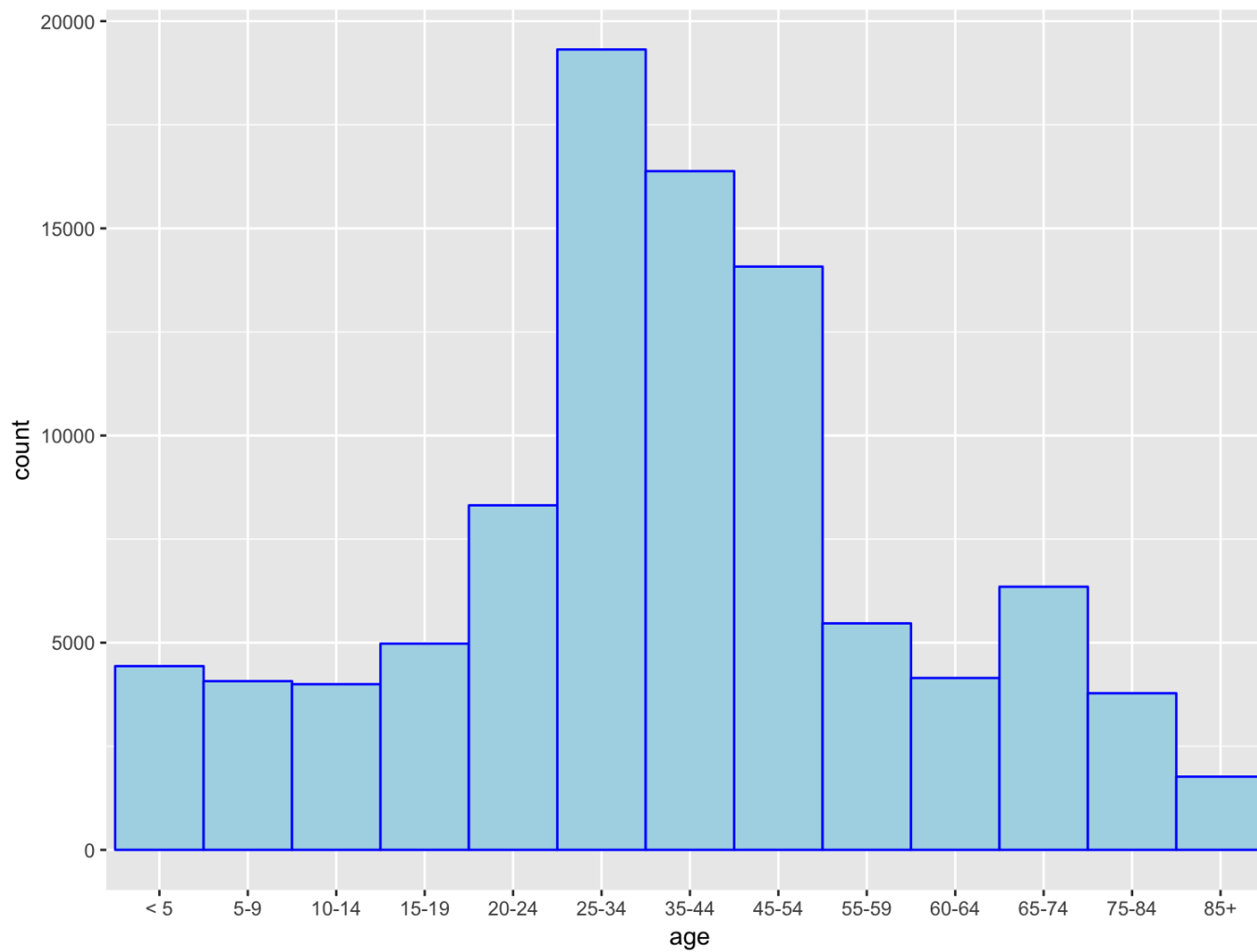


Count, Relative Frequency, Density

Bin	Count	Relative Frequency	Density
300-400	1	.056	.00056
400-500	4	.22	.0022
500-600	5	.28	.0028
600-700	4	.22	.0022
700-800	4	.22	.0022

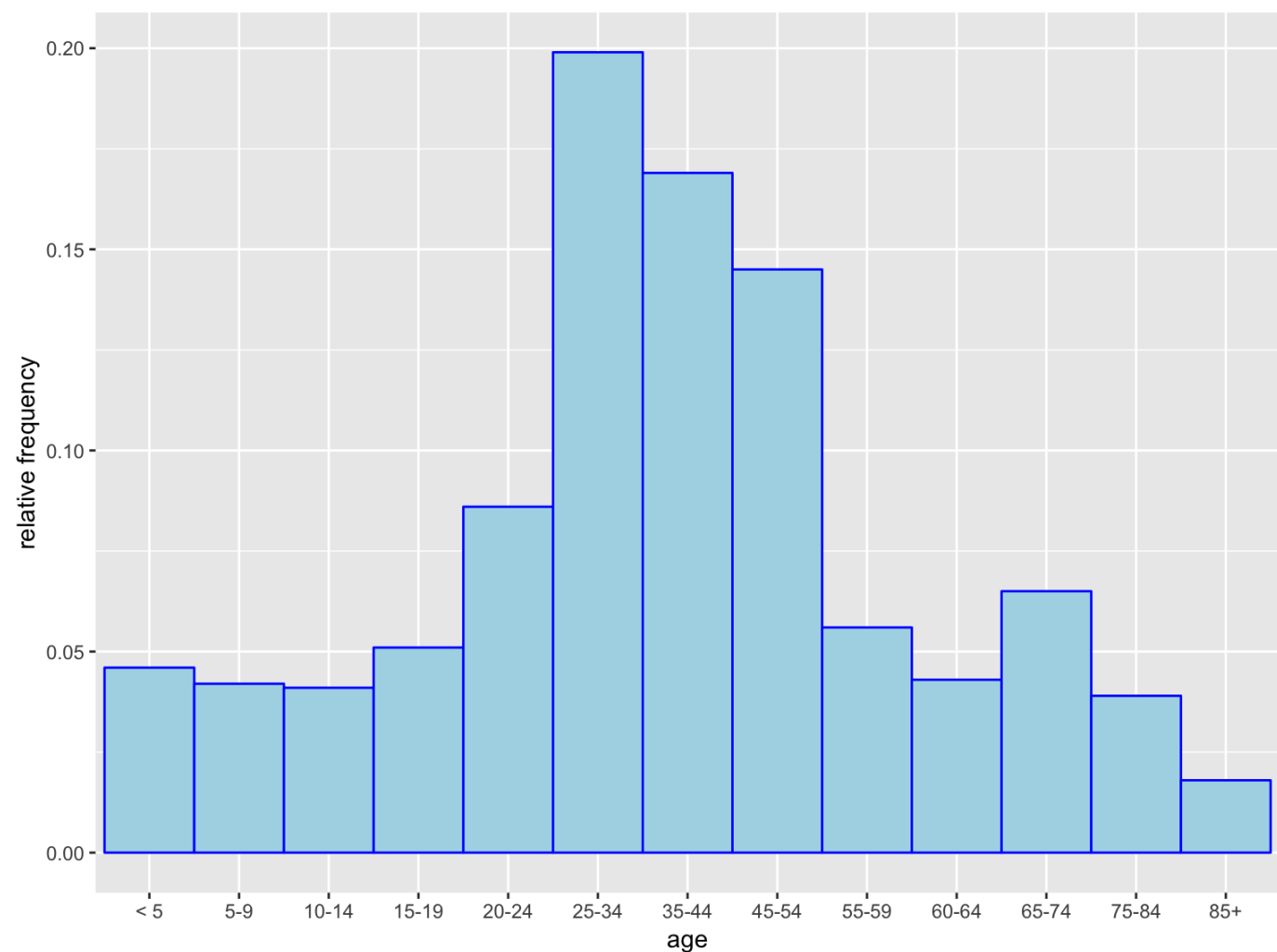
- How is relative frequency calculated?
- How is density calculated?

What's wrong with this histogram?

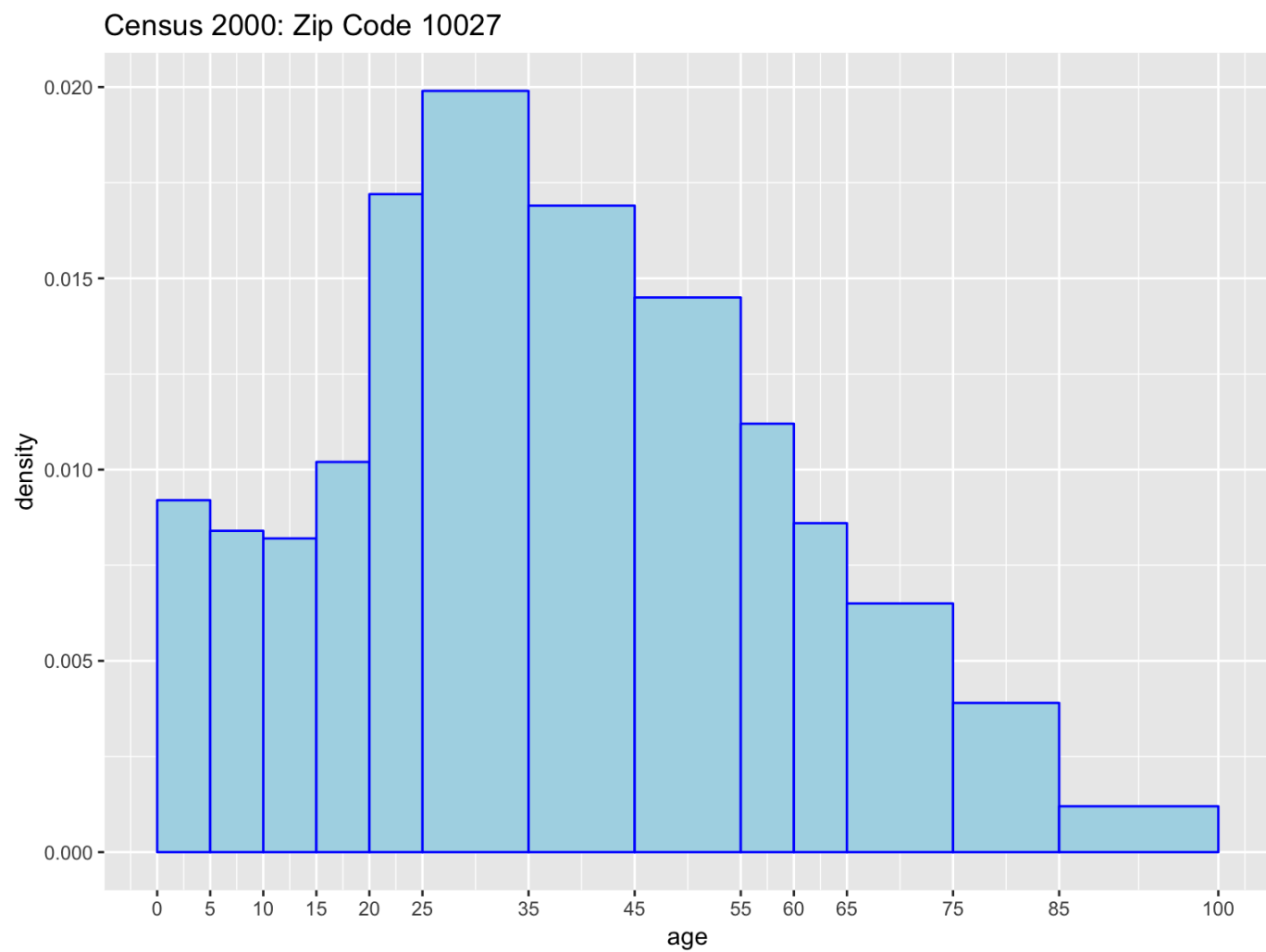


Relative frequency histogram

(still wrong...)



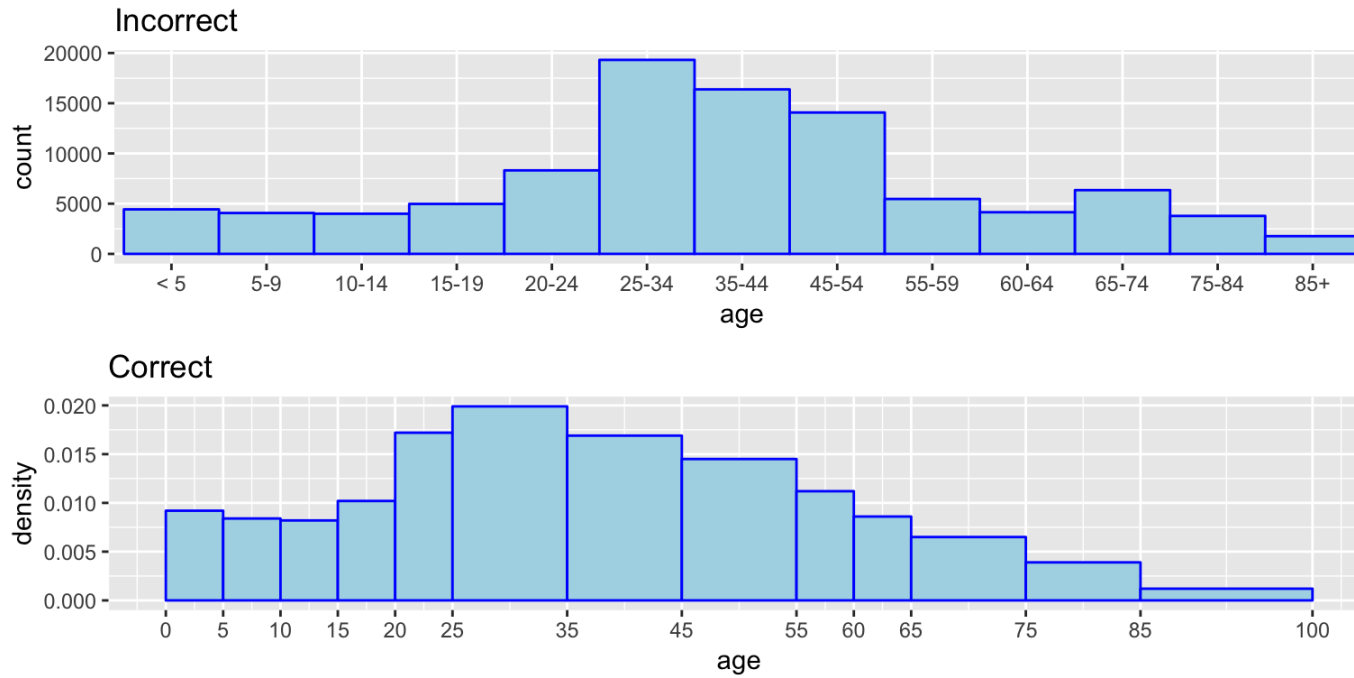
Density histogram with unequal bin (or class) widths



Density = RelFreq / Binwidth

Class	Frequency	RelFreq	ClassWidth	Density
< 5	4435	0.046	5	0.009
5-9	4072	0.042	5	0.008
10-14	3999	0.041	5	0.008
15-19	4977	0.051	5	0.010
20-24	8316	0.086	5	0.017
25-34	19317	0.199	10	0.020
35-44	16380	0.169	10	0.017
45-54	14077	0.145	10	0.014
55-59	5467	0.056	5	0.011
60-64	4148	0.043	5	0.009
65-74	6350	0.065	10	0.007
75-84	3781	0.039	10	0.004
85+	1767	0.018	15	0.001

Compare the histograms

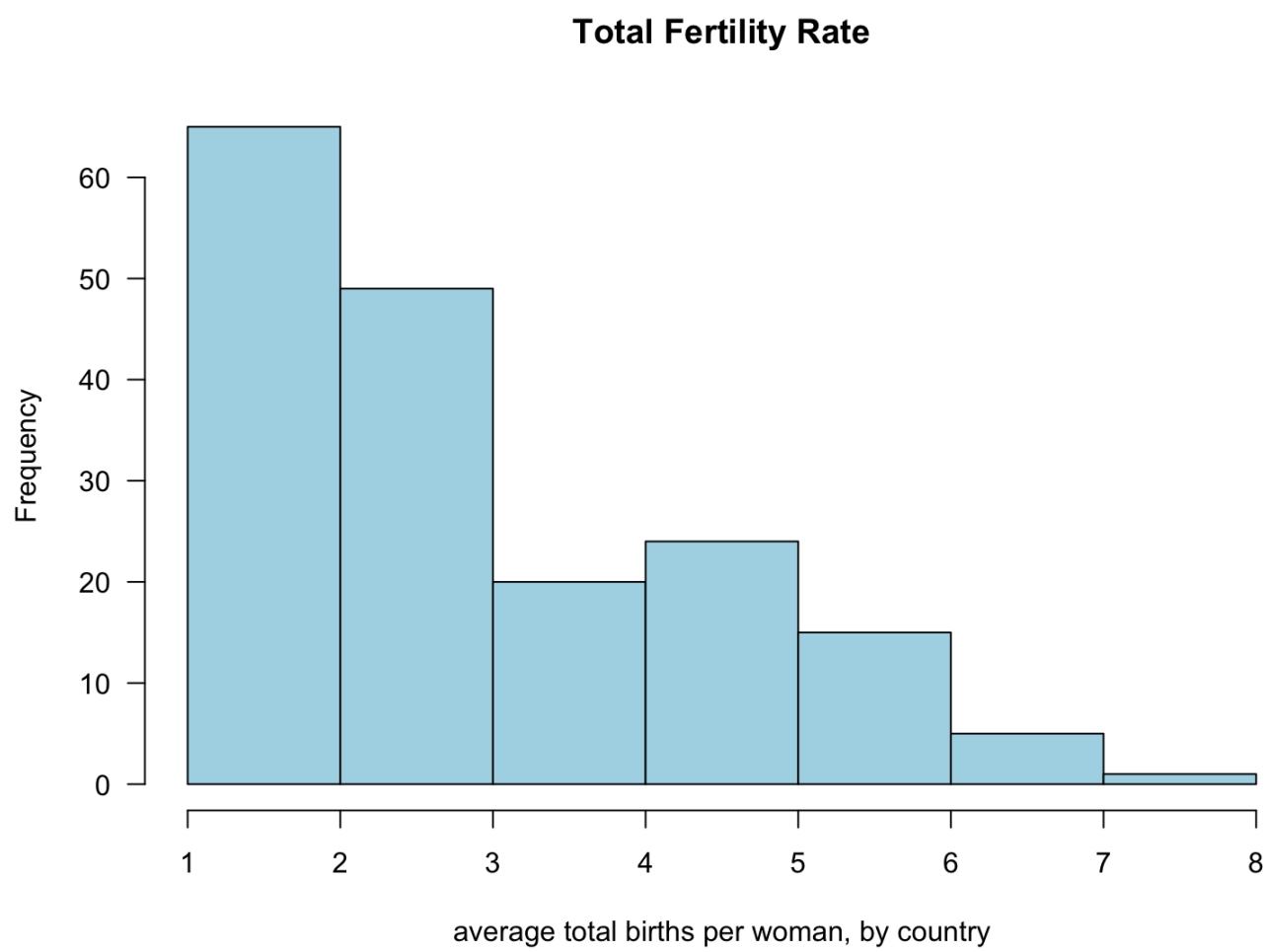


Source: <https://factfinder.census.gov/>

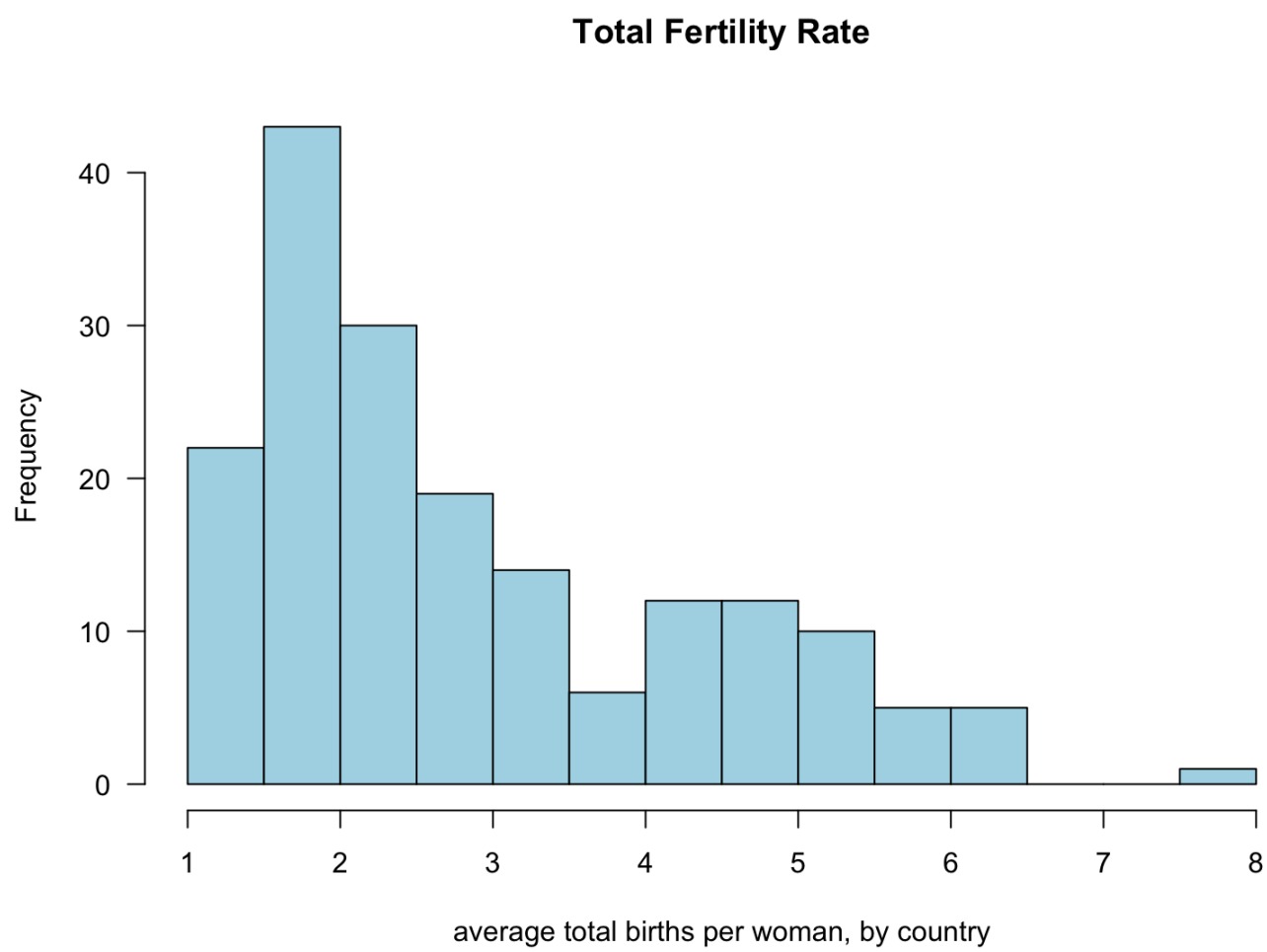
Binwidth

`'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.`

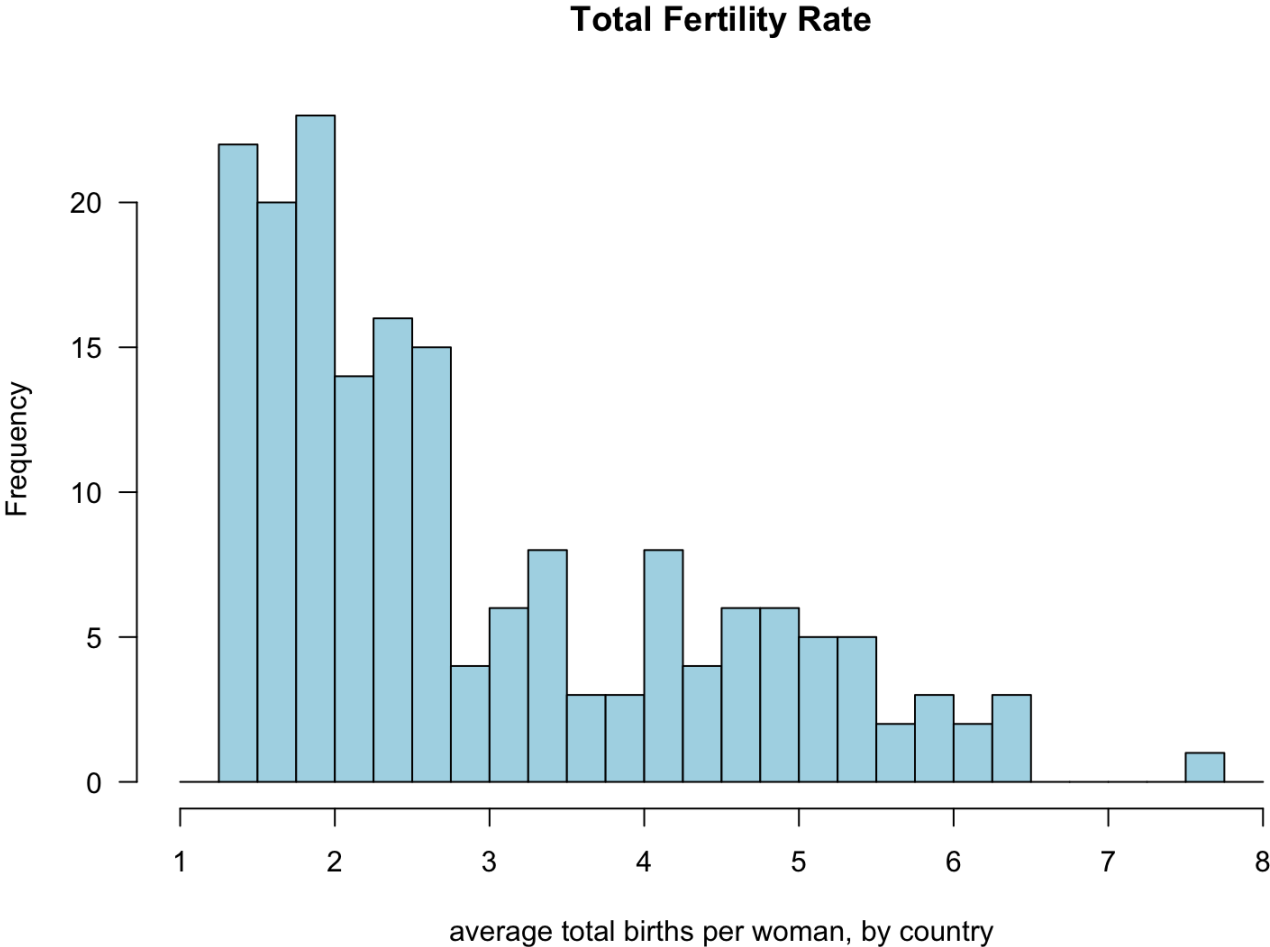
Histograms



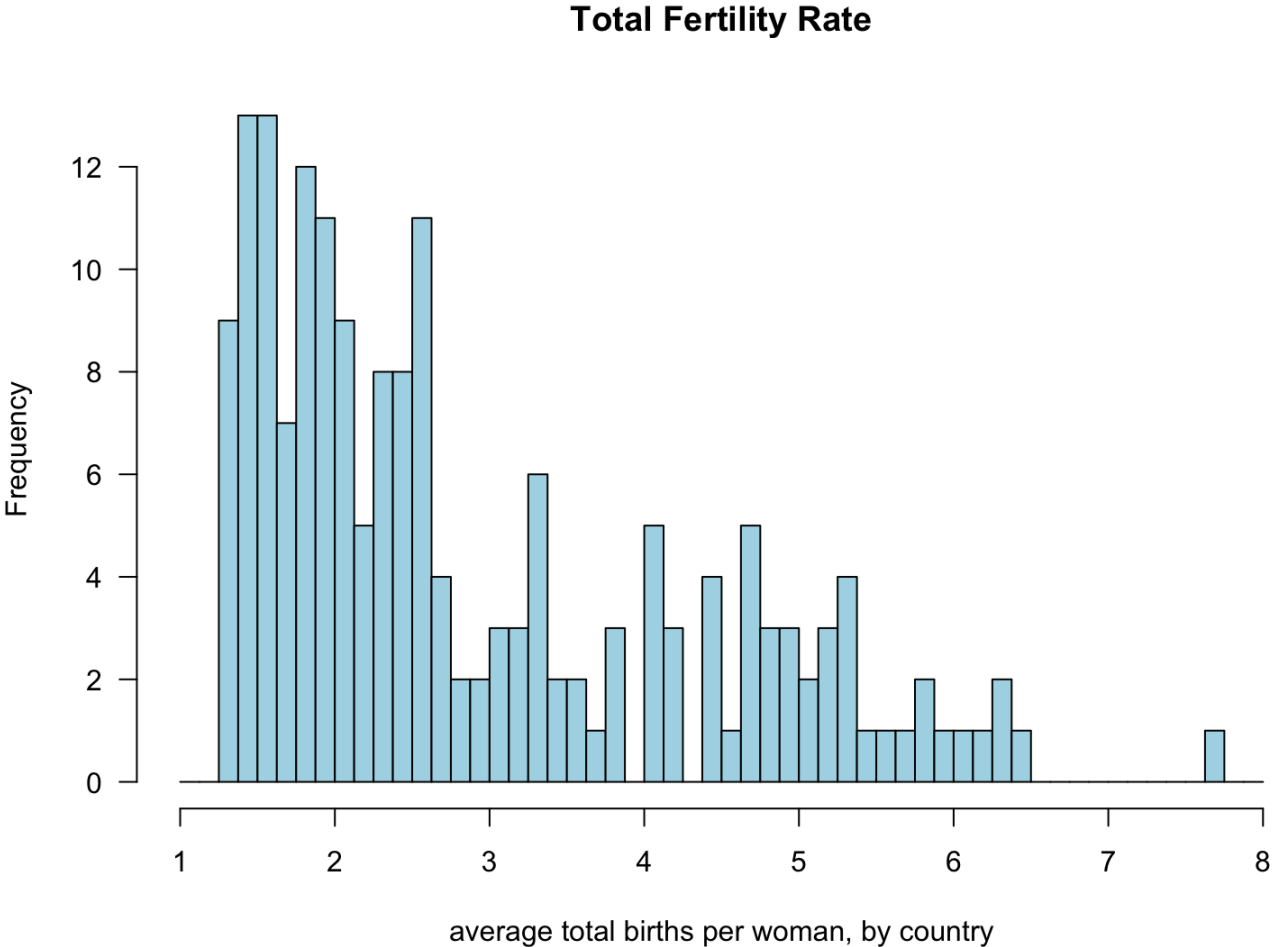
Histograms



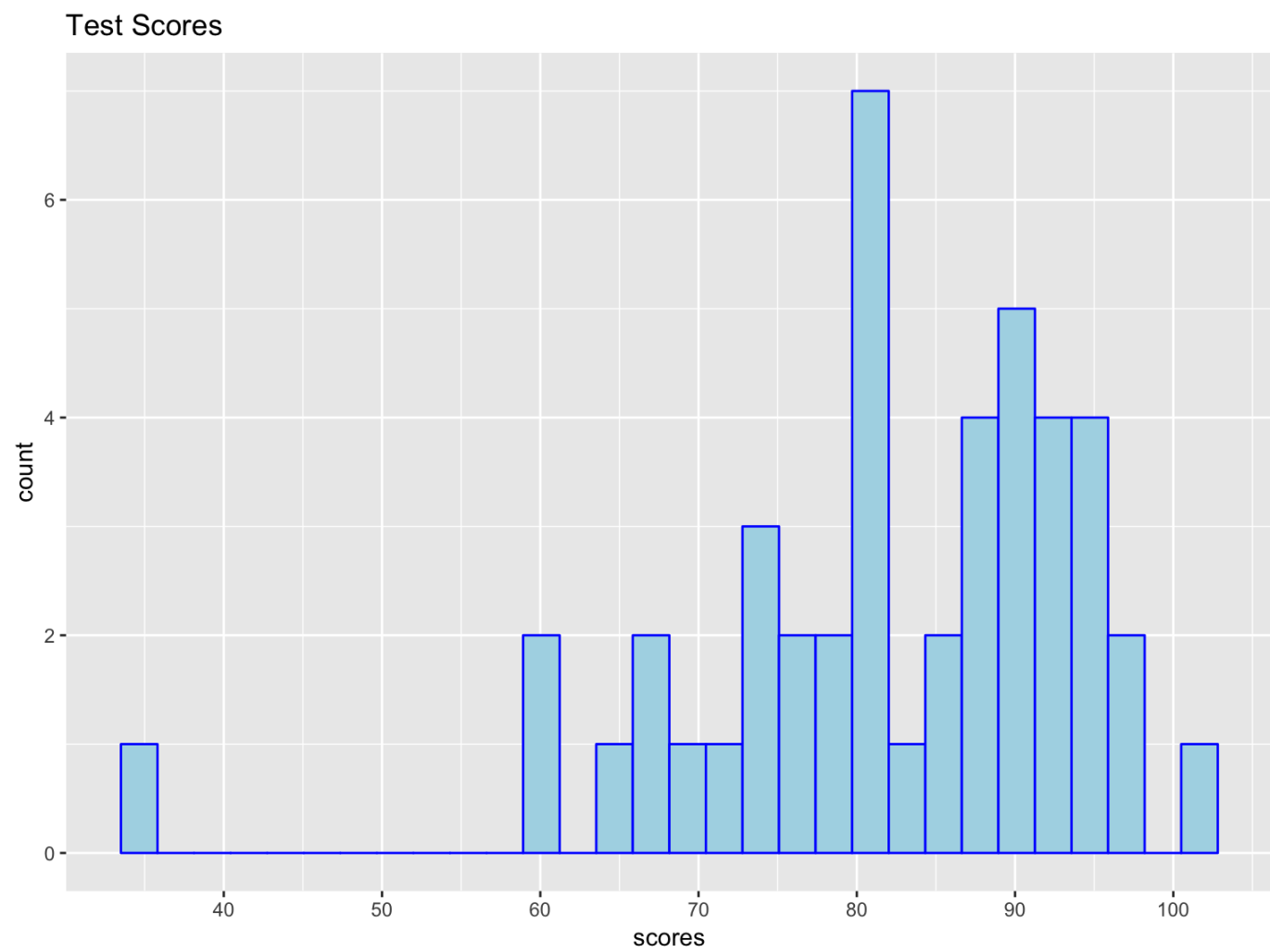
Histograms



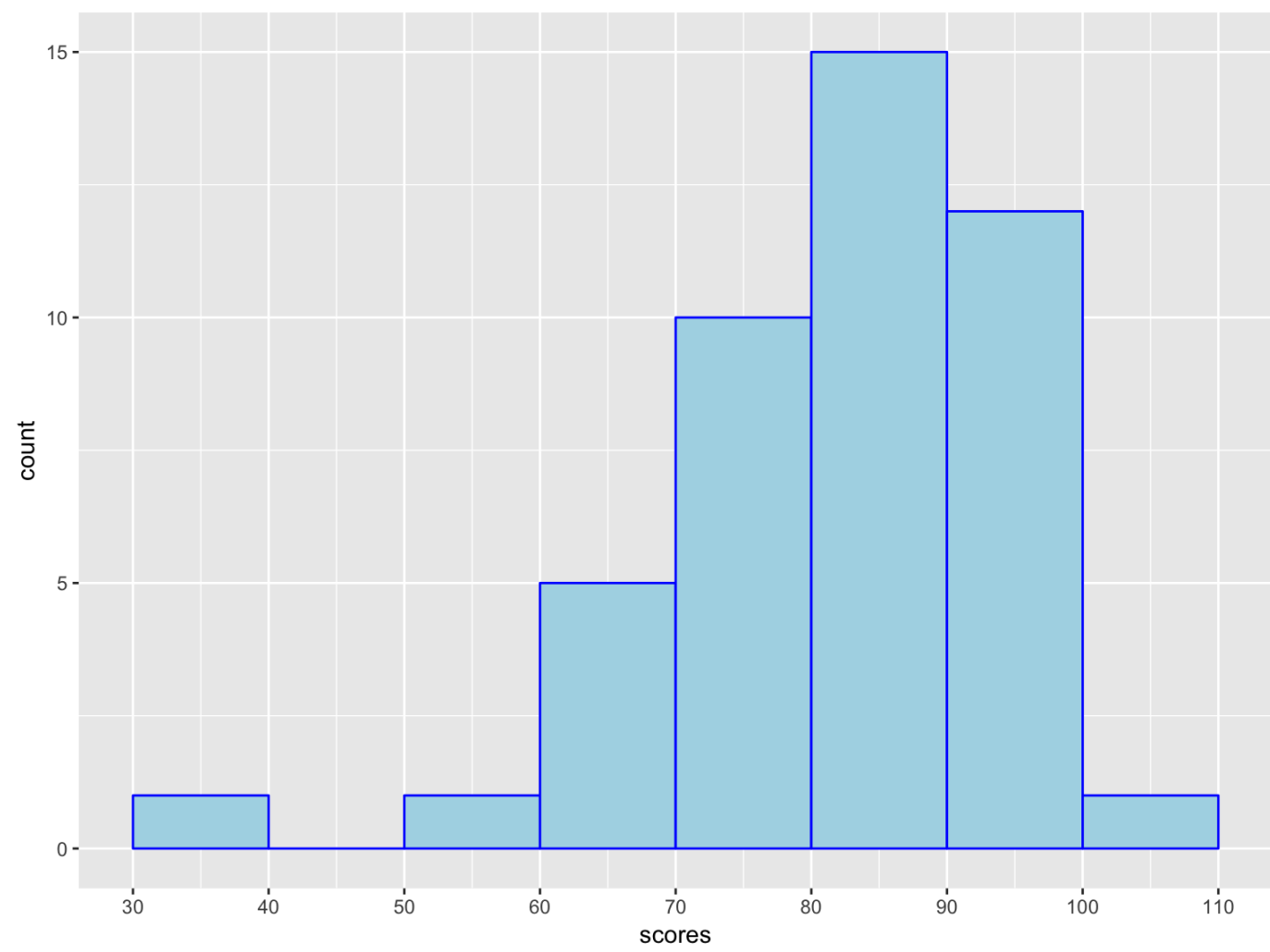
Histograms



Histograms



Fewer bins



Weekly Savings

Search:

	A	B	C	D
1	\$0.91	-\$95.25	-\$95.25	-\$195.25
2	\$1.50	-\$75.12	-\$75.12	-\$175.12
3	\$2.02	-\$61.77	-\$61.77	-\$161.77
4	\$2.40	-\$46.18	-\$46.18	-\$146.18
5	\$3.27	-\$39.82	-\$39.82	-\$139.82
6	\$4.77	-\$37.62	-\$37.62	-\$137.62
7	\$5.58	-\$22.62	-\$22.62	-\$122.62
8	\$6.65	-\$16.22	-\$16.22	-\$116.22
9	\$7.93	-\$6.19	-\$6.19	-\$106.19
10	\$10.86	-\$4.29	-\$4.29	-\$104.29
11	\$12.04	-\$3.25	-\$3.25	-\$103.25
12	\$13.92	-\$2.04	-\$2.04	-\$102.04
13	\$14.07	-\$1.52	-\$1.52	-\$101.52
14	\$14.23	\$0.58	\$0.58	-\$99.42

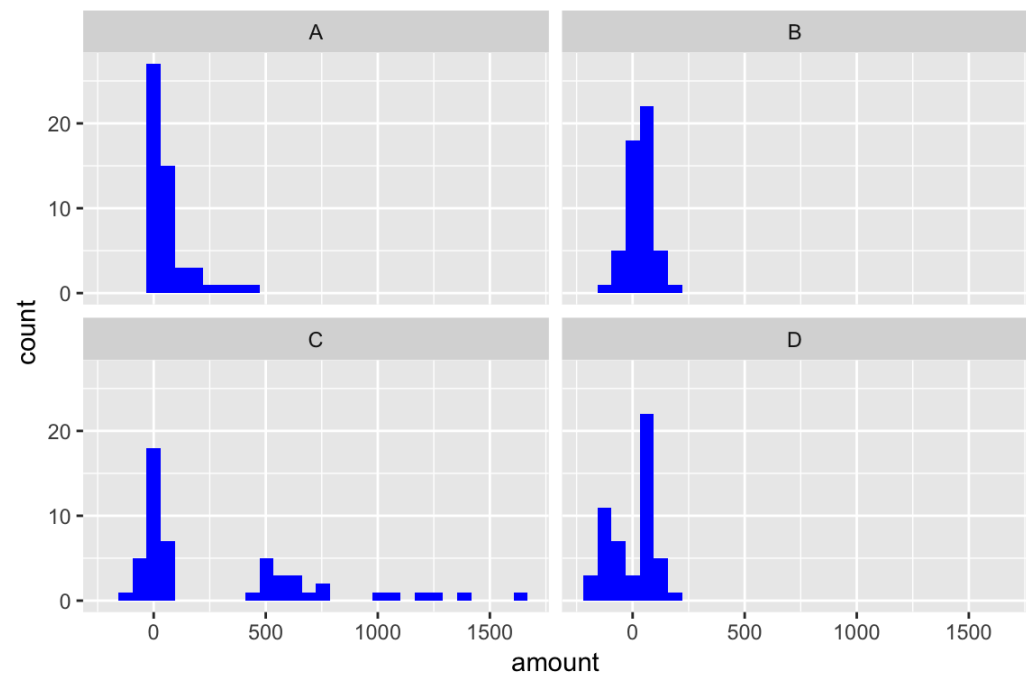
	A	B	C	D
15	\$14.58	\$8.38	\$8.38	-\$91.62
16	\$16.23	\$10.08	\$10.08	-\$89.92
17	\$18.85	\$13.60	\$13.60	-\$86.40
18	\$19.98	\$16.91	\$16.91	-\$83.09
19	\$24.44	\$17.47	\$17.47	-\$82.53
20	\$25.11	\$18.65	\$18.65	-\$81.35
21	\$25.68	\$20.10	\$20.10	-\$79.90
22	\$25.87	\$24.33	\$24.33	\$24.33
23	\$26.00	\$28.20	\$28.20	\$28.20
24	\$28.54	\$31.10	\$31.10	\$31.10
25	\$29.54	\$31.81	\$31.81	\$31.81
26	\$30.48	\$32.74	\$32.74	\$32.74
27	\$30.65	\$35.03	\$35.03	\$35.03
28	\$39.09	\$37.77	\$37.77	\$37.77
29	\$40.21	\$40.51	\$40.51	\$40.51
30	\$47.27	\$40.71	\$40.71	\$40.71
31	\$51.40	\$41.00	\$41.00	\$41.00

	A	B	C	D
32	\$52.31	\$45.79	\$457.93	\$45.79
33	\$57.08	\$48.48	\$484.75	\$48.48
34	\$58.27	\$49.30	\$493.00	\$49.30
35	\$65.17	\$49.78	\$497.84	\$49.78
36	\$65.55	\$52.18	\$521.84	\$52.18
37	\$73.49	\$52.62	\$526.19	\$52.62
38	\$73.73	\$54.15	\$541.53	\$54.15
39	\$74.93	\$55.68	\$556.83	\$55.68
40	\$82.54	\$59.80	\$597.98	\$59.80
41	\$85.92	\$62.60	\$626.02	\$62.60
42	\$92.27	\$65.14	\$651.41	\$65.14
43	\$95.69	\$65.37	\$653.71	\$65.37
44	\$104.58	\$70.36	\$703.56	\$70.36
45	\$124.60	\$76.70	\$766.99	\$76.70
46	\$192.96	\$78.21	\$782.15	\$78.21
47	\$194.34	\$103.50	\$1,035.00	\$103.50
48	\$199.99	\$109.22	\$1,092.22	\$109.22

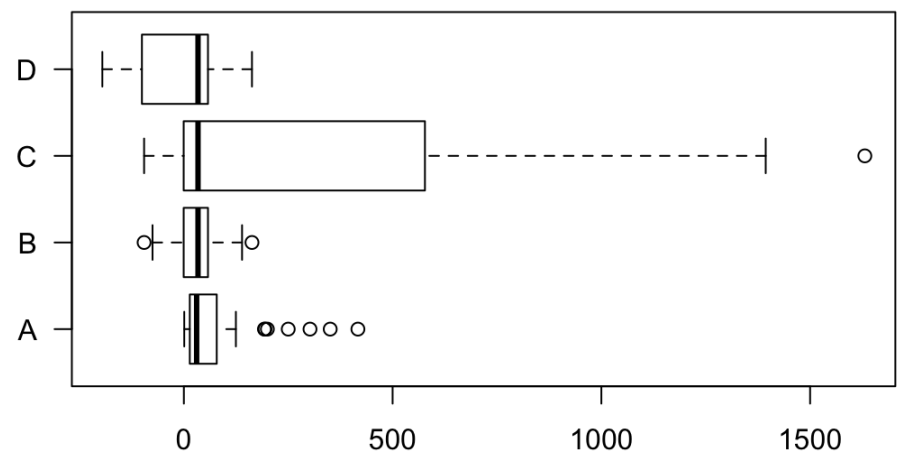
	A	B	C	D
49	\$249.96	\$119.50	\$1,194.99	\$119.50
50	\$302.12	\$128.15	\$1,281.47	\$128.15
51	\$350.54	\$139.37	\$1,393.66	\$139.37
52	\$416.85	\$163.11	\$1,631.09	\$163.11

Showing 1 to 52 of 52 entries

Histograms

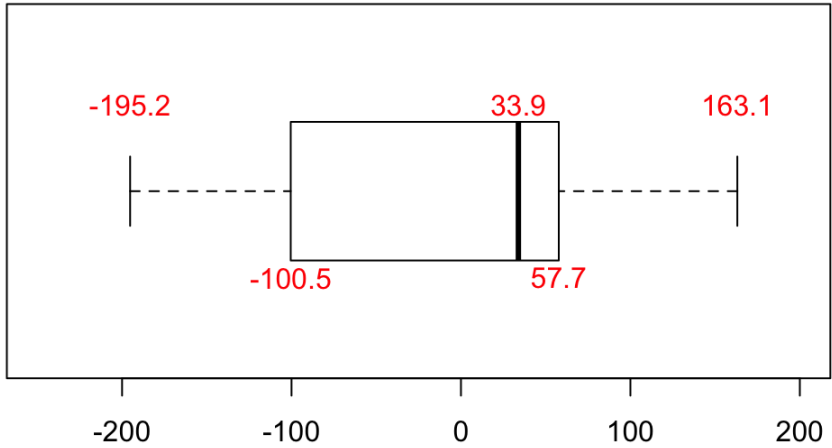


Boxplots



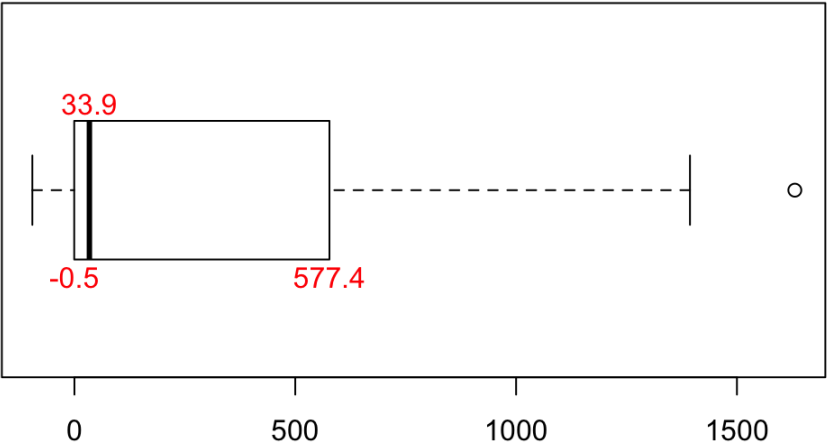
Boxplot (Person “D”)

##	min	lower-hinge	median	upper-hinge	max
##	-195.2	-100.5	33.9	57.7	163.1

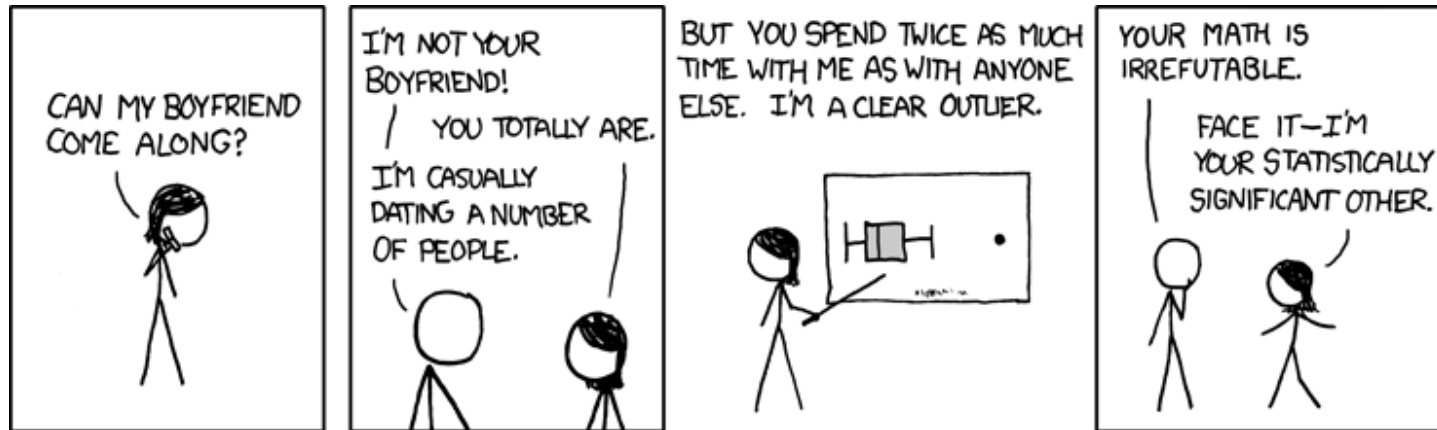


Boxplot with outliers (Person “C”)

##	min	lower-hinge	median	upper-hinge	max
##	-95.249	-0.473	33.889	577.408	1631.089

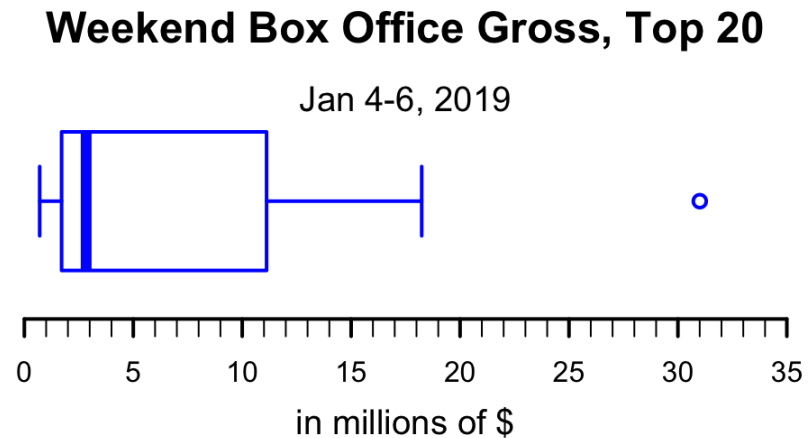


What does it take to be an outlier?



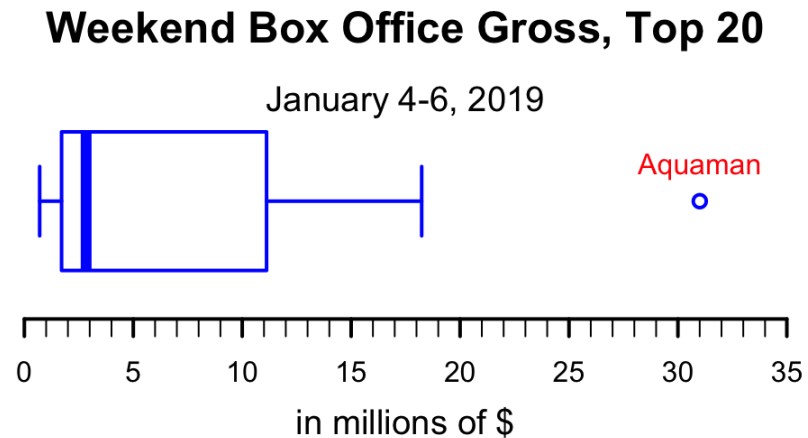
https://www.explainxkcd.com/wiki/index.php/539:_Boyfriend

What does it take to be an outlier?



Source: <http://www.boxofficemojo.com/weekend/chart/>

What does it take to be an outlier?

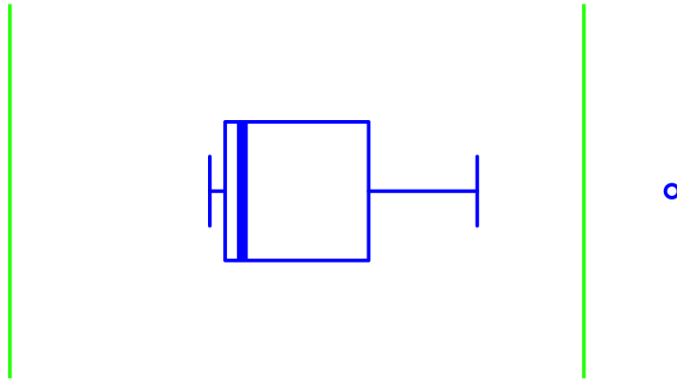


fences:

$1.5 \times$ hinge spread above upper-hinge

$1.5 \times$ hinge spread below lower-hinge

Fences

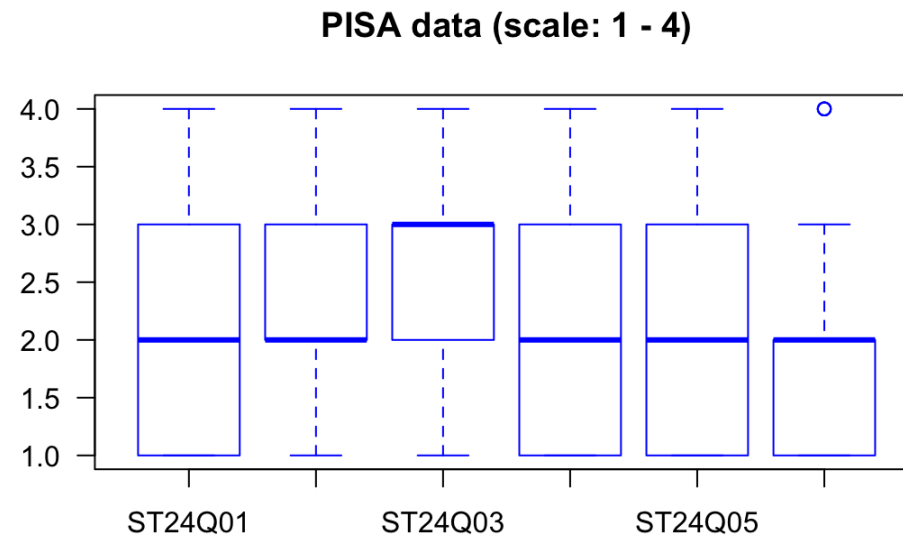


fences:

$1.5 \times$ hinge spread above upper-hinge

$1.5 \times$ hinge spread below lower-hinge

Not for discrete data

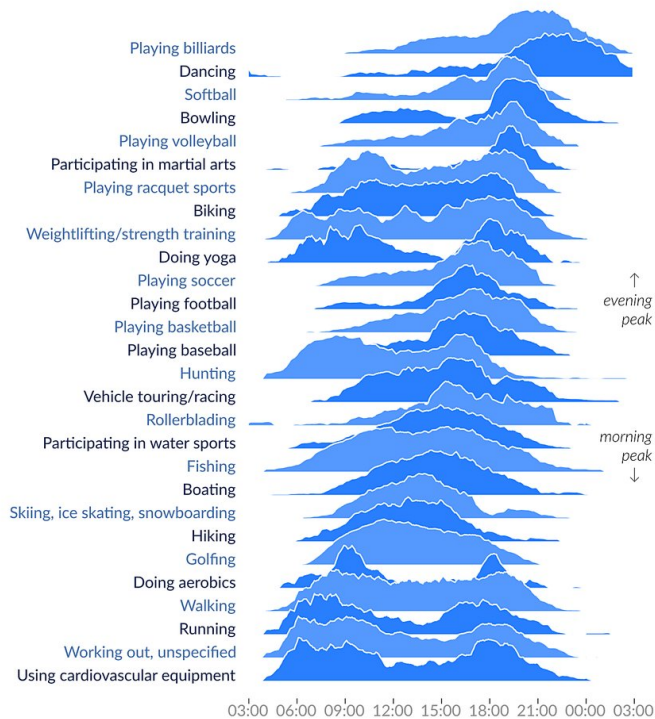


Source: R likert::pisaitems dataset

Ridgeline plot

Peak time of day for sports and leisure

Number of participants throughout the day compared to peak popularity.
Note the morning-and-evening everyday workouts, the midday hobbies,
and the evenings/late nights out.



@hnrklnbrg | Source: American Time Use Survey

Source: <https://eagereyes.org/blog/2017/joy-plots>

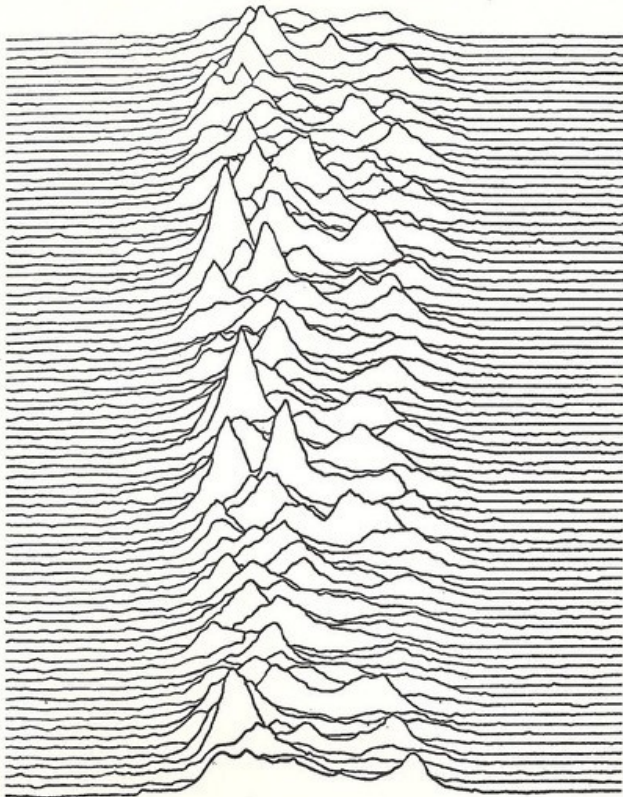
Additional resources:

<http://blog.revolutionanalytics.com/2017/07/joyplots.html>

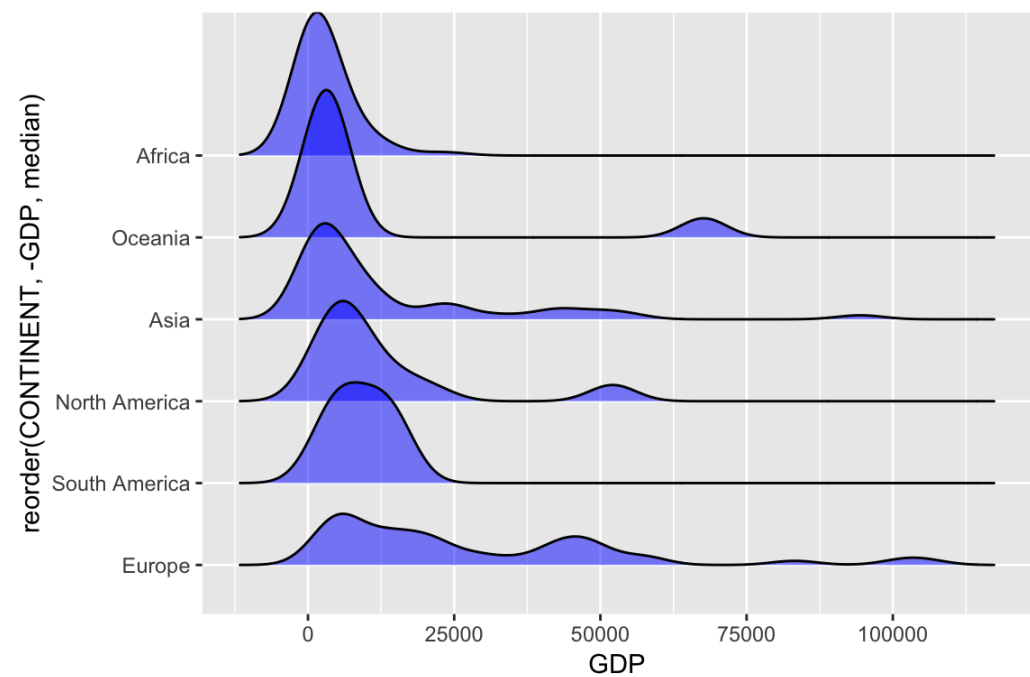
Ridgeline plot inspiration

Jocelyn Bell discovers first radio pulsars, 1967

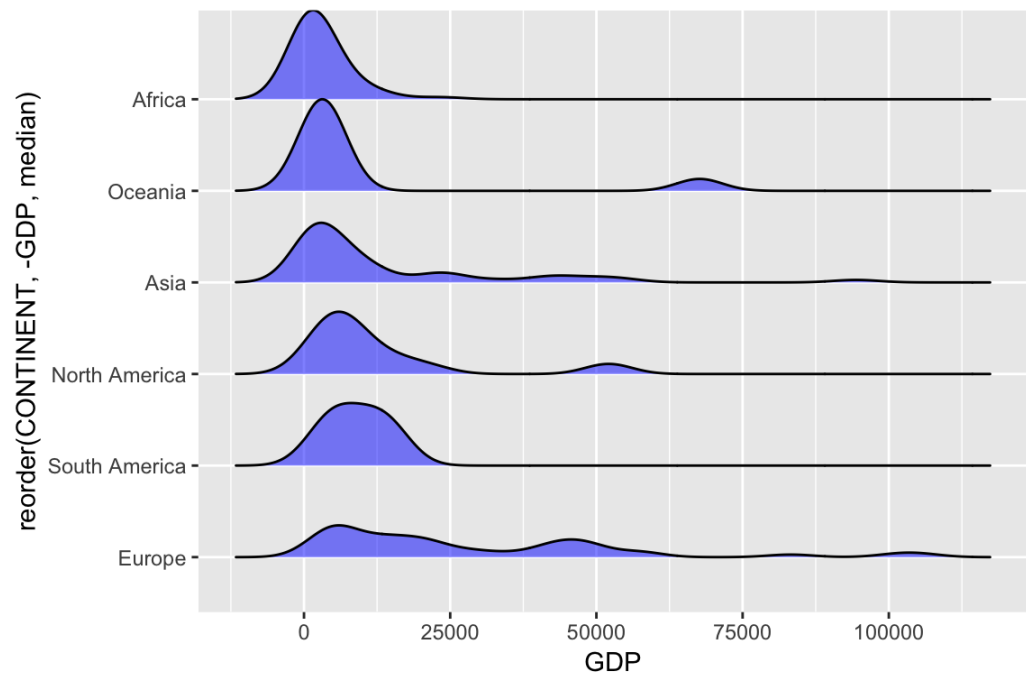
6.7: Successive pulses from the first pulsar discovered, CP 1919, are here superimposed vertically. The pulses occur every 1.337 seconds. They are caused by a rapidly-spinning neutron star.



Ridgeline plot



Ridgeline plot, change scale



Ridgeline vs. boxplot



Source: <https://twitter.com/lenkiefier/status/916823350726610946>

ggribges package

CRAN <https://CRAN.R-project.org/package=ggribges>

Github <https://github.com/clauswilke/ggribges>

Package vignette(s) <https://cran.r-project.org/web/packages/ggribges/vignettes/introduction.html>

<https://cran.r-project.org/web/packages/ggribges/vignettes/gallery.html>

Package manual <https://cran.r-project.org/web/packages/ggribges/ggribges.pdf>