Continuous Variables, pt. 2

Weekly Savings

Search:

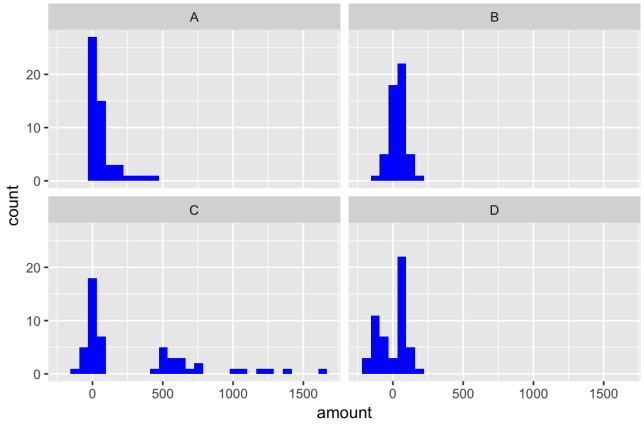
	A	B	C	D
1	$0.91	-$95.25	-$95.25	-$195.25
2	$1.50	-$75.12	-$75.12	-$175.12
3	$2.02	-$61.77	-$61.77	-$161.77
4	$2.40	-$46.18	-$46.18	-$146.18
5	$3.27	-$39.82	-$39.82	-$139.82
6	$4.77	-$37.62	-$37.62	-$137.62
7	$5.58	-$22.62	-$22.62	-$122.62
8	$6.65	-$16.22	-$16.22	-$116.22
9	$7.93	-$6.19	-$6.19	-$106.19
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
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

	A	B	C	D
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
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
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

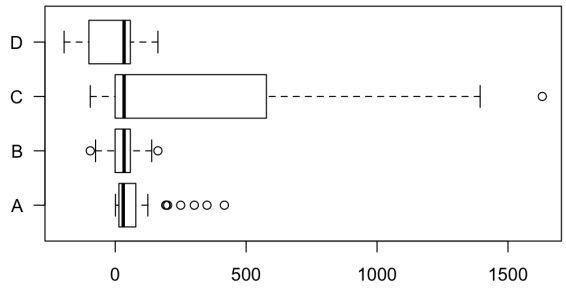
	A	B	C	D
52	$416.85	$163.11	$1,631.09	$163.11

Showing 1 to 52 of 52 entries

Histograms

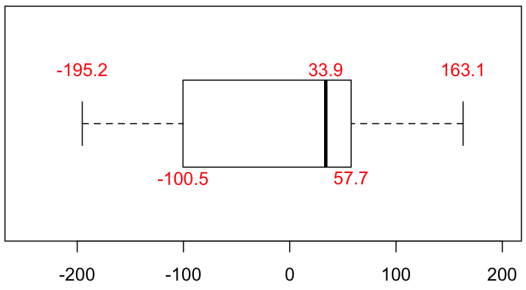


Boxplots



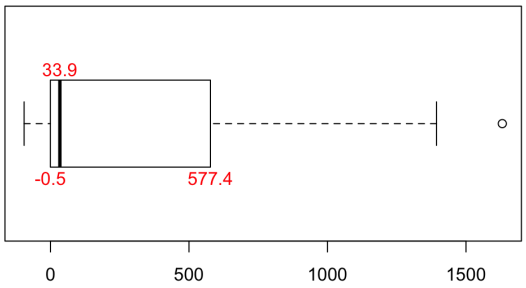
Boxplot (Person “D”)

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



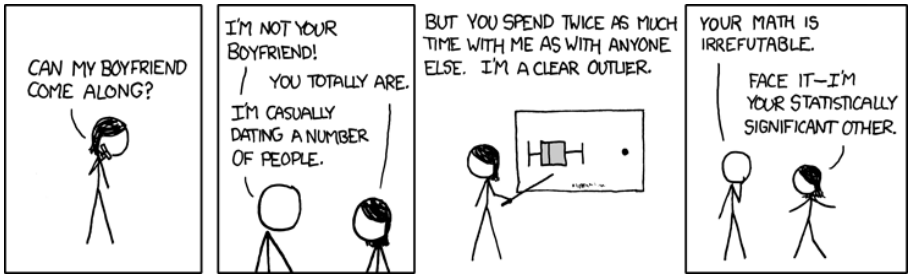
Boxplot with outliers (Person “C”)

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



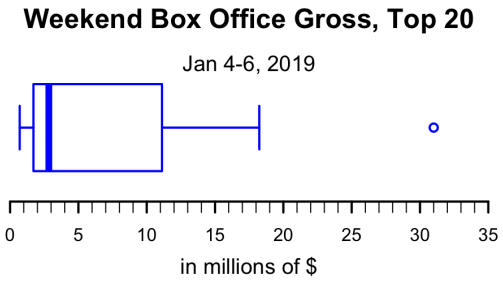
What does it take to be an outlier?

What does it take to be an outlier?



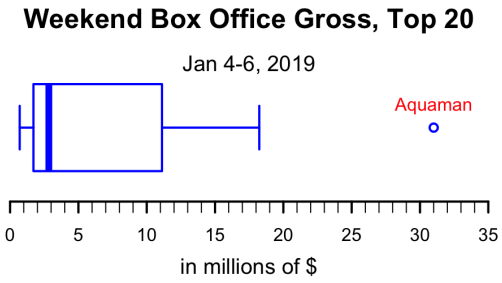
https://www.explainxkcd.com/wiki/index.php/539:_Boyfriend

What does it take to be an outlier?

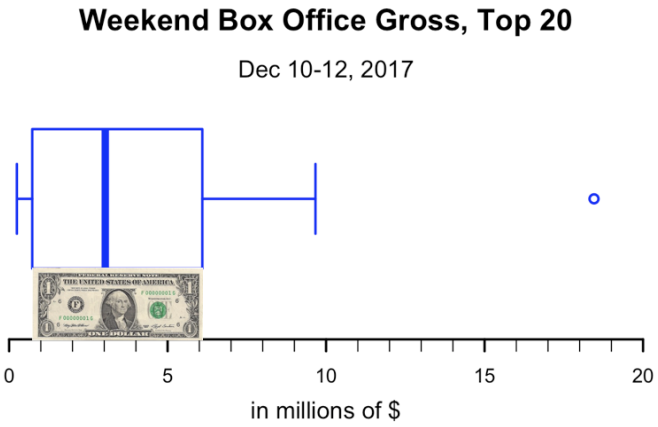


Source: http://www.boxofficemojo.com/weekend/chart/

What does it take to be an outlier?

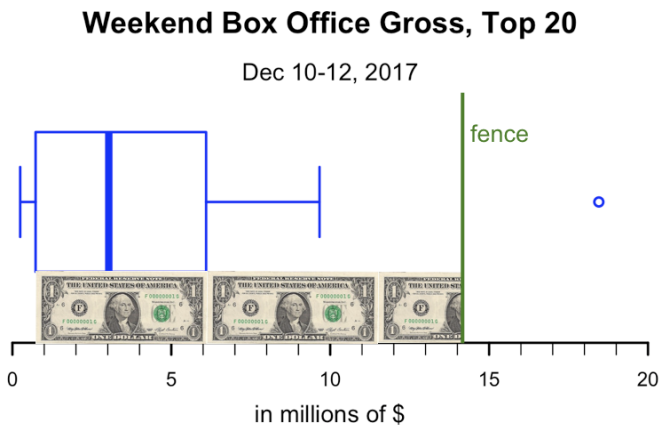


What does it take to be an outlier?



"H-spread" or fourth spread (upper hinge - lower hinge)

What does it take to be an outlier?

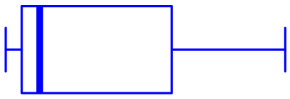


fences:

1.5 x hinge spread above upper-hinge

1.5 x hinge spread below lower-hinge

Fences

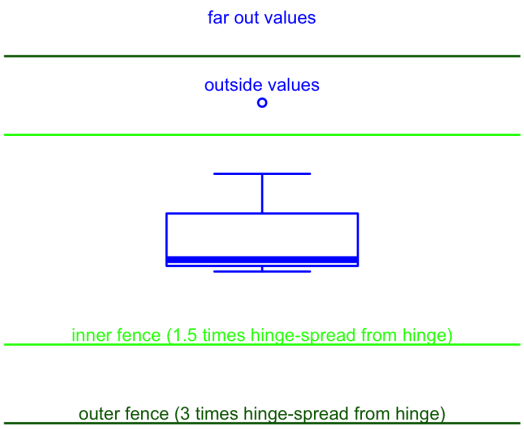


fences:

1.5 x hinge spread above upper-hinge

1.5 x hinge spread below lower-hinge

Tukey’s original boxplot



Quartiles

```
boxoffice
```

```
##  [1]  0.703  0.923  1.005  1.168  1.609  1.808  1.843  1.903  2.147  2.368
## [11]  3.303  4.674  4.755  5.735  9.110 13.127 13.203 15.861 18.238 31.003
```

```
fivenum(boxoffice) %>% set_names(fivenumnames)
```

```
##         min lower-hinge      median upper-hinge         max
##       0.703       1.709       2.835      11.118      31.003
```

```
quantile(boxoffice)
```

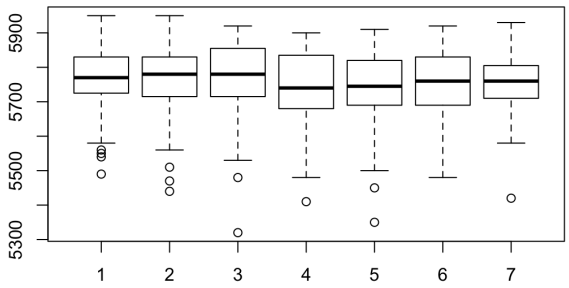
```
##     0%    25%    50%    75%   100%
##  0.703  1.758  2.835 10.114 31.003
```

See: ?quantile for different methods

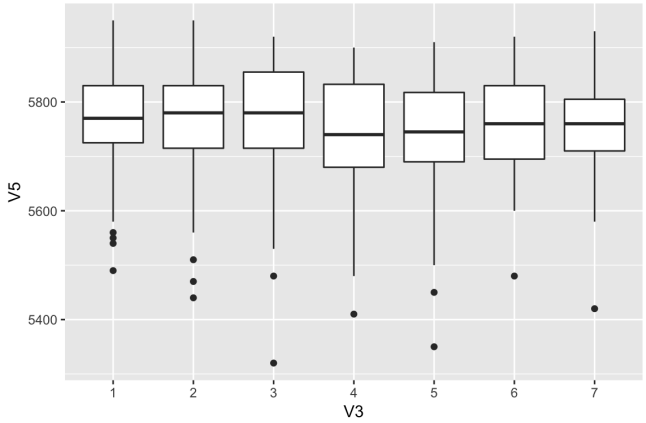
Sometimes boxplots are drawn using the IQR (interquartile range) instead of hinge spread

base R vs. ggplot2

```
library(mlbench)
data(Ozone)
boxplot(V5 ~ V3, data = Ozone)
```




```
ggplot(Ozone, aes(V3, V5)) + geom_boxplot()
```



Box plot stats

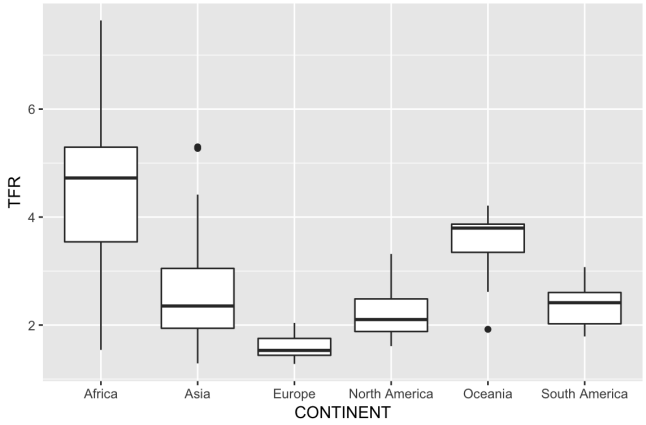
```r
# base R
boxplot.stats(df$`Weekend Gross`)
```

```
## $stats
## [1]  0.703  1.709  2.835 11.118 18.238
##
## $n
## [1] 20
##
## $conf
## [1] -0.489  6.160
##
## $out
## [1] 31
```

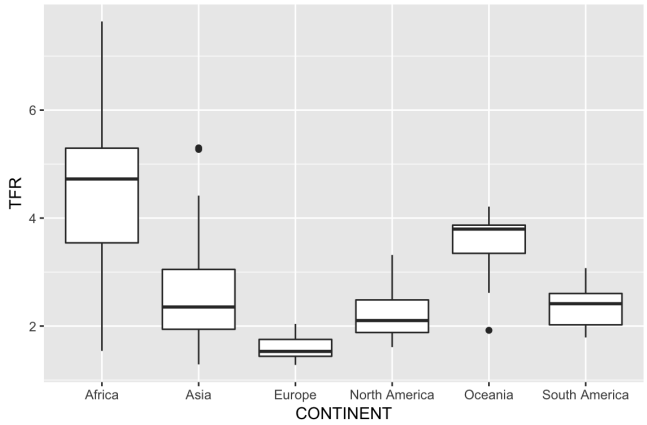
```r
# ggplot2
g <- ggplot(df, aes(1, `Weekend Gross`)) + geom_boxplot()
ggplot_build(g)$data[[1]]
```

ymin	lower	middle	upper	ymax	outliers	notchupper	notchlow
0.703	1.76	2.83	10.1	18.2	31.00328	5.79	-0.1

Multiple box plots

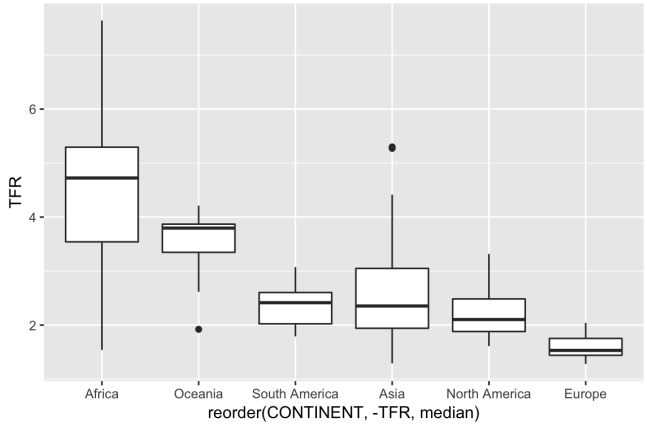


Multiple box plots

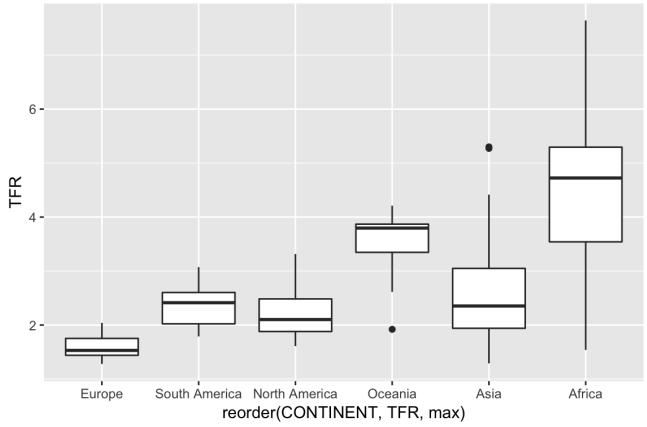


COUNTRY	CONTINENT	TFR
Afghanistan	Asia	5.27
Timor-Leste	Asia	5.30
COUNTRY	**CONTINENT**	**TFR**
Australia	Oceania	1.92

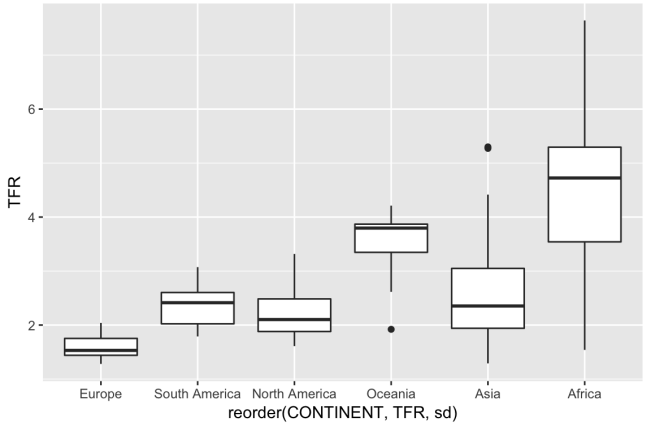
Reorder by median



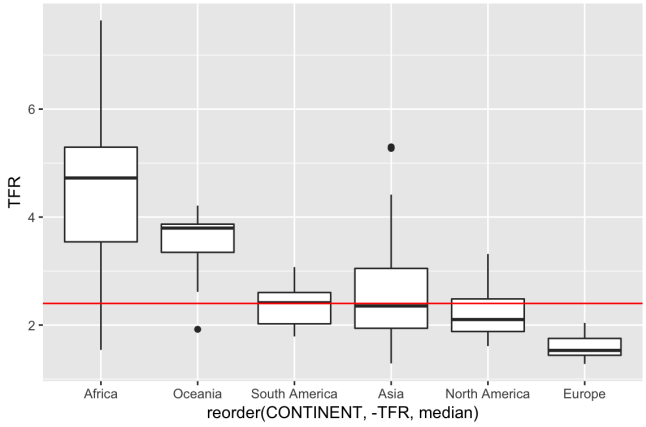
Reorder by maximum value



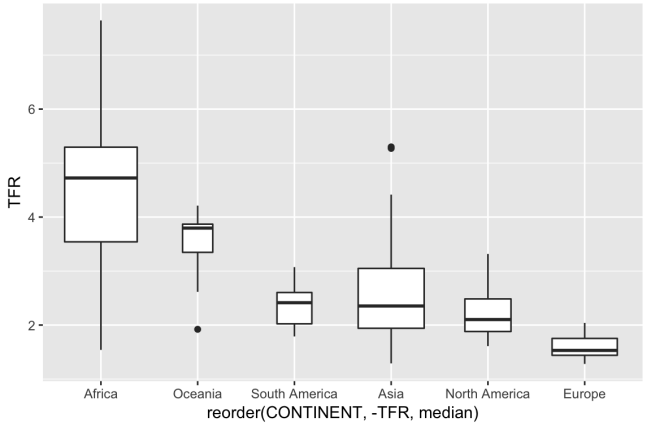
Reorder by standard deviation



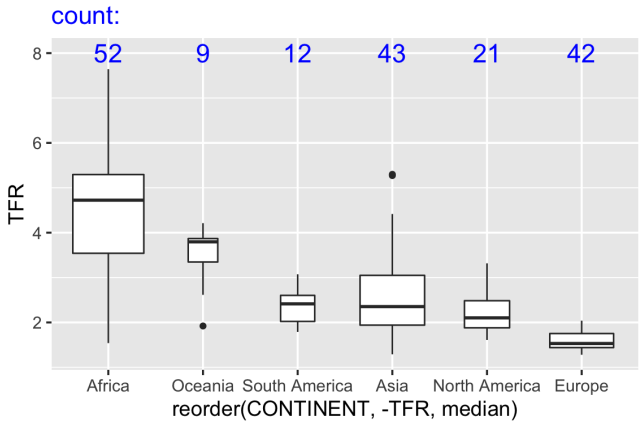
Add overall median line



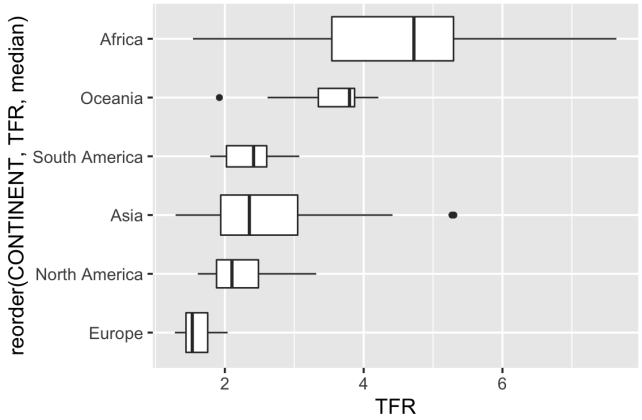
Variable width box plots



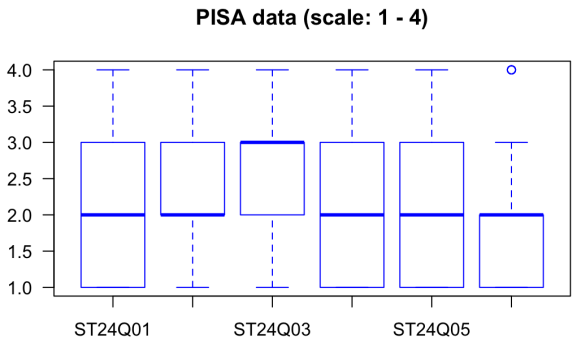
Add continent country count



Horizontal boxplot

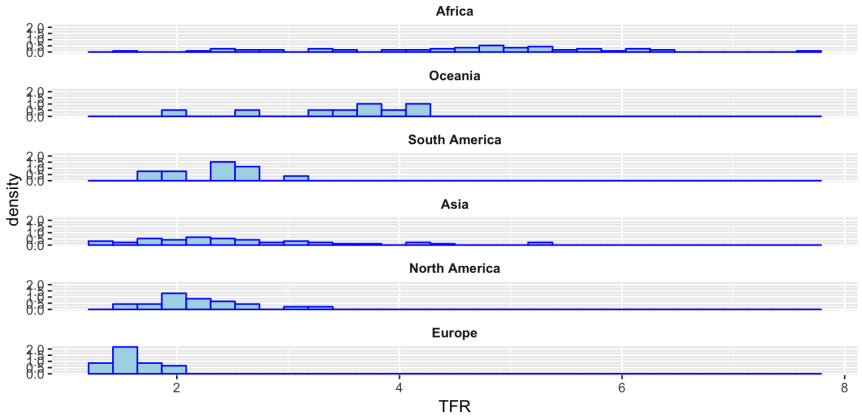


Not for discrete data

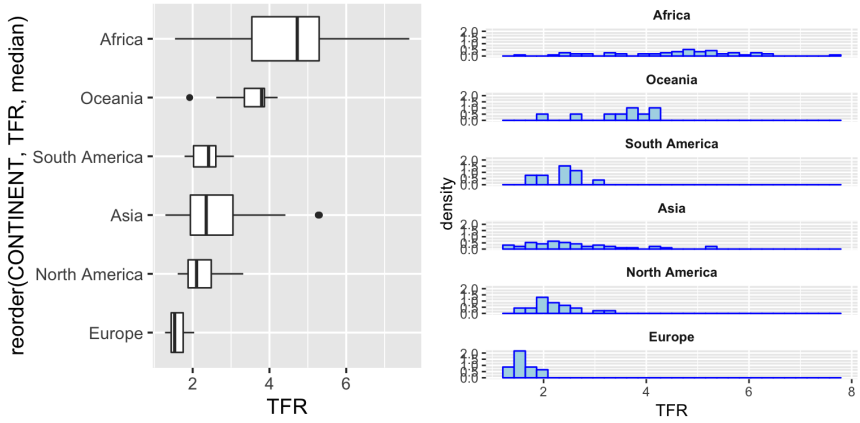


Source: R likert::pisaitems dataset

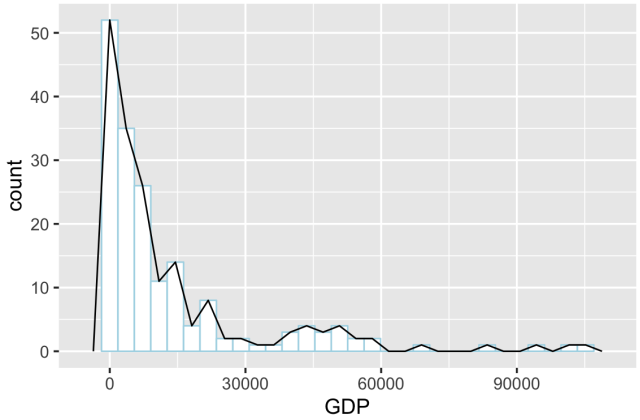
Multiple density histograms, ordered by median



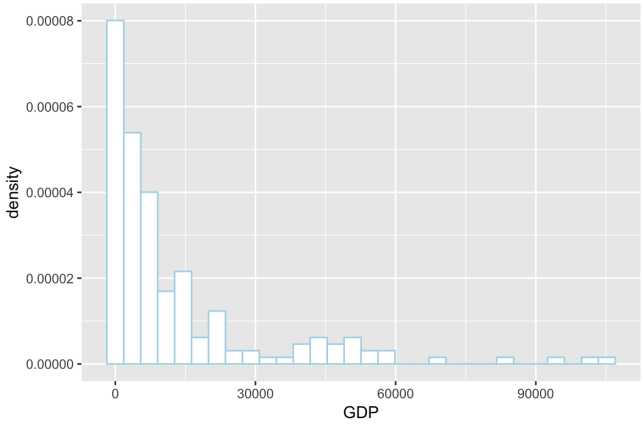
Boxplots vs. histograms



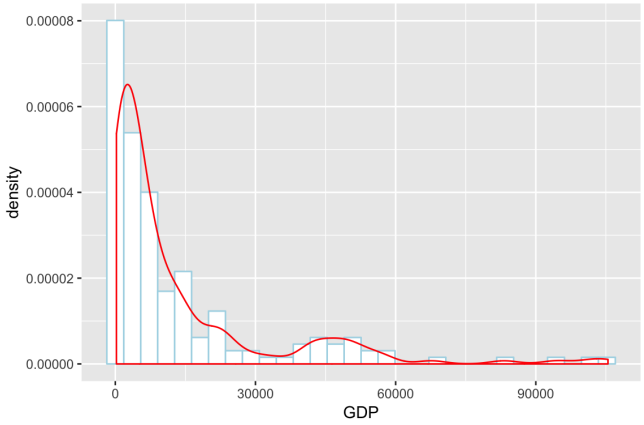
Frequency polygon



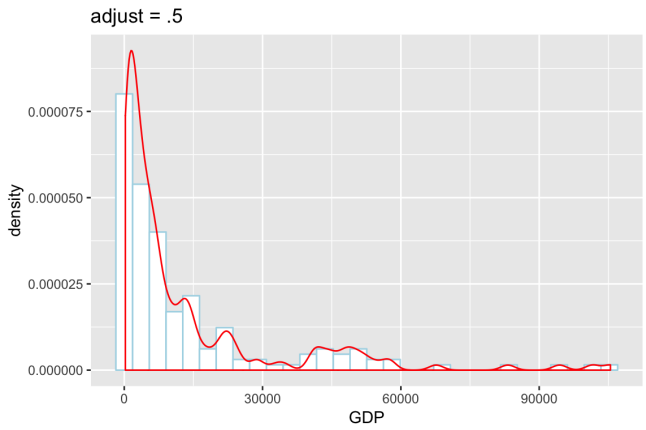
Density histogram



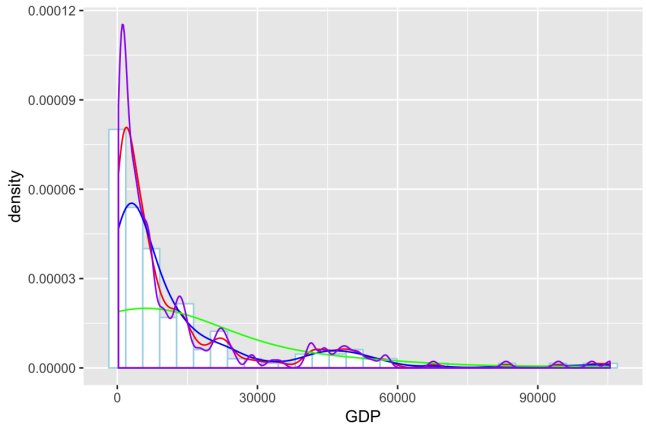
Density curve



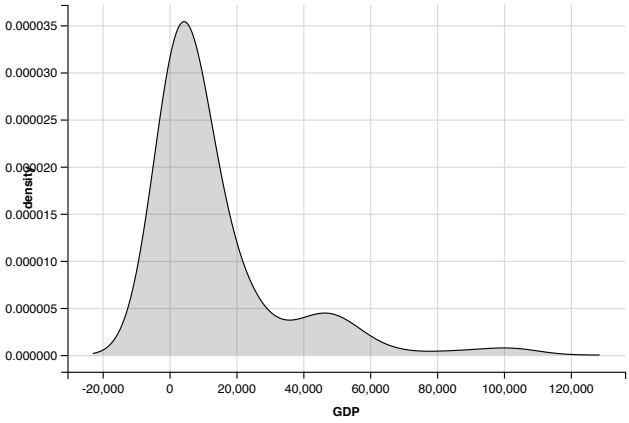
Density curve



Density curve: varying smoothing bandwidths

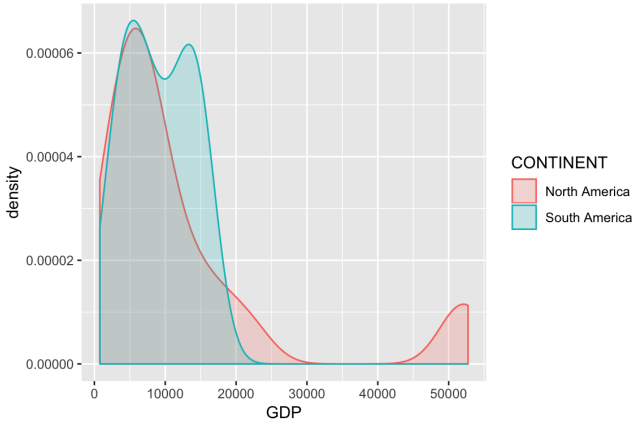


Density curve: varying smoothing bandwidths (`ggvis`)

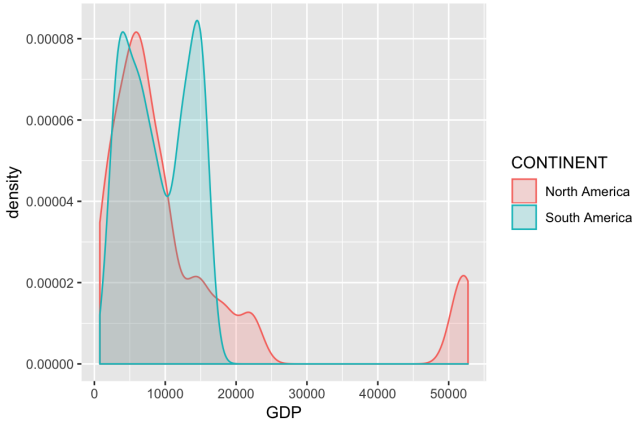


See also: http://ggvis.rstudio.com/0.1/quick-examples.html#histograms

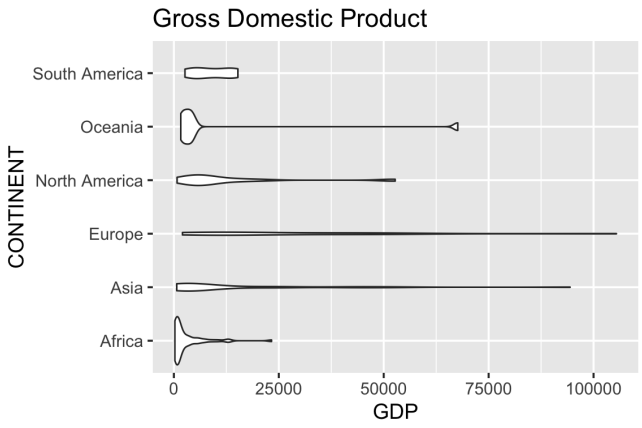
Density curves



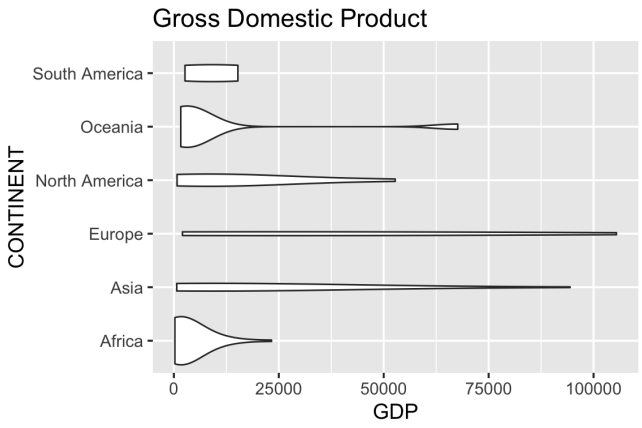
Density curves



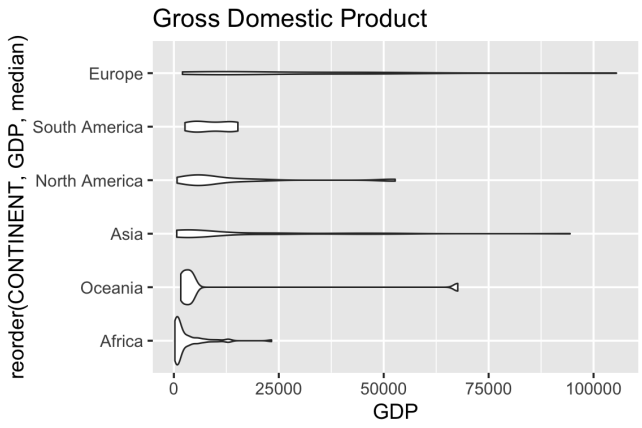
Violin plots



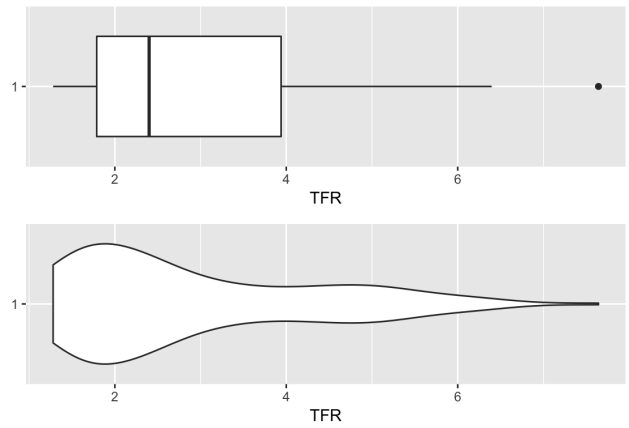
Violin plots, change bandwidth



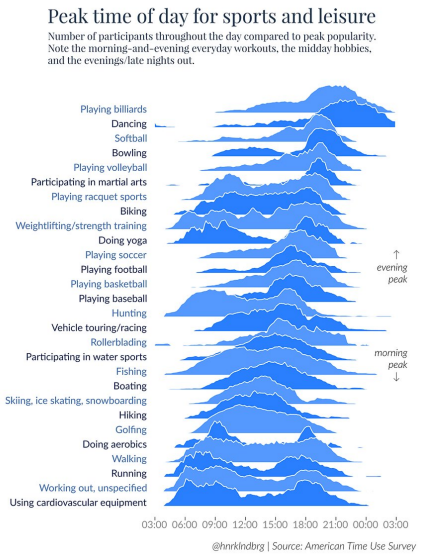
Violin plots, ordered by median



Box plot vs. violin plot



Ridgeline plot



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

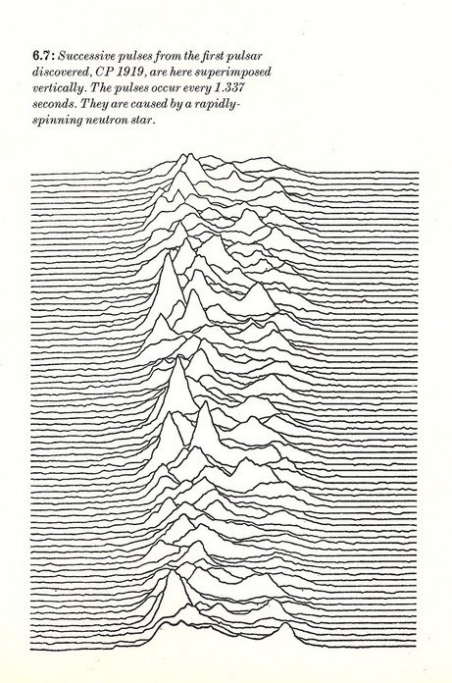
Additional resources:

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

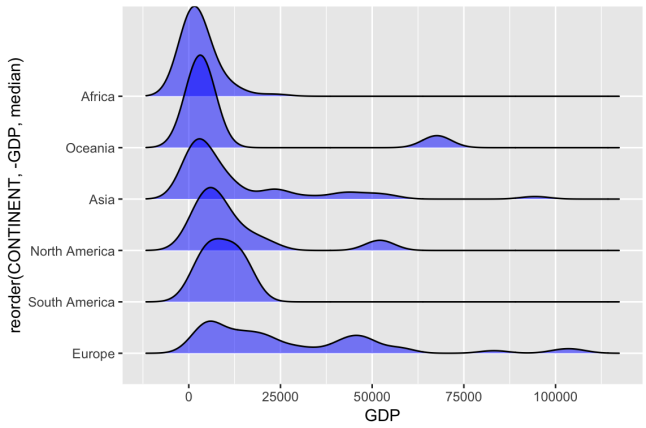
https://blogs.scientificamerican.com/sa-visual/pop-culture-pulsar-origin-story-of-joy-division-s-unknown-pleasures-album-cover-video/

Ridgeline plot inspiration

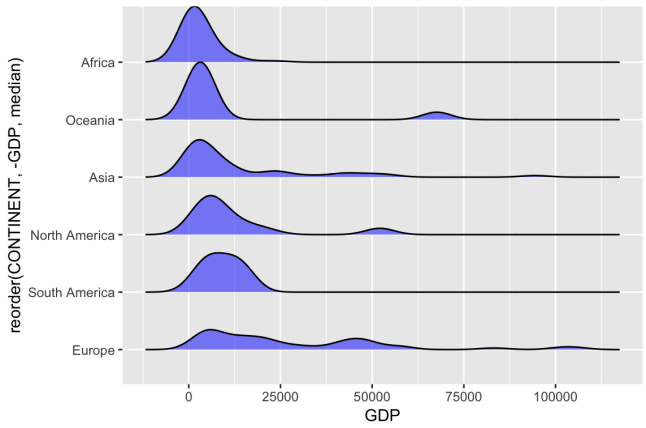
Jocelyn Bell discovers first radio pulsars, 1967



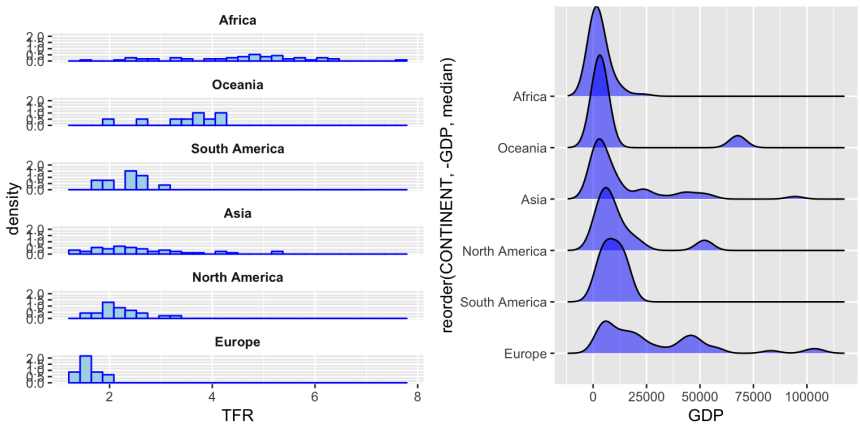
Ridgeline plot



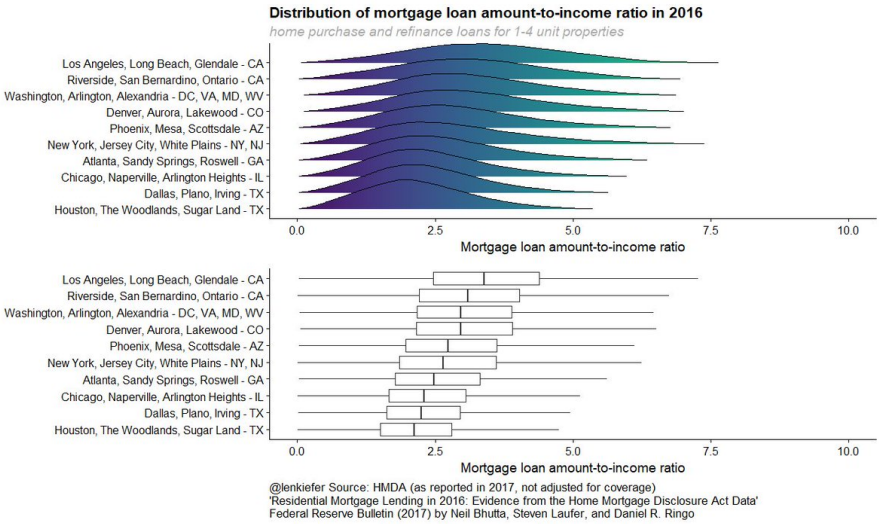
Ridgeline plot, change scale



Histogram vs. ridgeline



Ridgeline vs. boxplot



Source: https://twitter.com/lenkiefer/status/916823350726610946

`ggridge` package

CRAN https://CRAN.R-project.org/package=ggridges

Github https://github.com/clauswilke/ggridges

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

https://cran.r-project.org/web/packages/ggridges/vignettes/gallery.html

Package manual https://cran.r-project.org/web/packages/ggridges/ggridges.pdf