

# Package: lvplot (via r-universe)

September 5, 2024

**Title** Letter Value 'Boxplots'

**Version** 0.2.1.9000

**Description** Implements the letter value 'boxplot' which extends the standard 'boxplot' to deal with both larger and smaller number of data points by dynamically selecting the appropriate number of letter values to display.

**License** GPL (>= 2)

**Depends** R (>= 2.14)

**Imports** ggplot2 (>= 3.3.0), grid, RColorBrewer, tibble

**Suggests** depth

**Encoding** UTF-8

**LazyData** true

**LazyDataCompression** xz

**RoxygenNote** 7.2.1

**Repository** <https://hadley.r-universe.dev>

**RemoteUrl** <https://github.com/hadley/lvplot>

**RemoteRef** HEAD

**RemoteSha** a51f81ea29cdef13602442b2c806fcd6f9f21a32

## Contents

census . . . . .	2
determineDepth . . . . .	2
geom_lv . . . . .	3
LVboxplot . . . . .	6
lvtable . . . . .	8
ontime . . . . .	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

census	<i>County demographics based on 1980 US Census</i>
--------	--

---

**Description**

County level statistics based on the 1980 US Census.

**Usage**

census

**Format**

A data frame with 10 variables

**county** County name

**FIPS** FIPS county code

**Latitude,Longitude** Geographic location of county centers

**JanTmp,JulTmp** (normalized) Temperatures in January & July

**JanSun,JulSun** (normalized) Sunshine measurement in January & July

**Elevtn** Elevation above sea level

**totalpop** Population

---

determineDepth	<i>Determine depth of letter values needed for n observations.</i>
----------------	--

---

**Description**

Determine depth of letter values needed for n observations.

**Usage**

determineDepth(n, k = NULL, alpha = NULL, perc = NULL)

**Arguments**

n number of observation to be shown in the LV boxplot

k number of letter value statistics used

alpha if supplied, depth k is calculated such that  $(1-\alpha)100$  intervals of an LV statistic do not extend into neighboring LV statistics.

perc if supplied, depth k is adjusted such that perc percent outliers are shown

**Details**

Supply one of k, alpha or perc.

## Description

An extension of standard boxplots which draws k letter statistics. Conventional boxplots (Tukey 1977) are useful displays for conveying rough information about the central 50% of the data and the extent of the data. For moderate-sized data sets ( $n < 1000$ ), detailed estimates of tail behavior beyond the quartiles may not be trustworthy, so the information provided by boxplots is appropriately somewhat vague beyond the quartiles, and the expected number of “outliers” and “far-out” values for a Gaussian sample of size  $n$  is often less than 10 (Hoaglin, Iglewicz, and Tukey 1986). Large data sets ( $n \approx 10,000 - 100,000$ ) afford more precise estimates of quantiles in the tails beyond the quartiles and also can be expected to present a large number of “outliers” (about  $0.4 + 0.007n$ ). The letter-value box plot addresses both these shortcomings: it conveys more detailed information in the tails using letter values, only out to the depths where the letter values are reliable estimates of their corresponding quantiles (corresponding to tail areas of roughly  $2^{-i}$ ); “outliers” are defined as a function of the most extreme letter value shown. All aspects shown on the letter-value boxplot are actual observations, thus remaining faithful to the principles that governed Tukey’s original boxplot.

## Usage

```
geom_lv(  
  mapping = NULL,  
  data = NULL,  
  stat = "lv",  
  position = "dodge",  
  outlier.colour = "black",  
  outlier.shape = 19,  
  outlier.size = 1.5,  
  outlier.stroke = 0.5,  
  na.rm = TRUE,  
  varwidth = FALSE,  
  width.method = "linear",  
  show.legend = NA,  
  inherit.aes = TRUE,  
  ...  
)
```

GeomLv

scale\_fill\_lv(...)

```
stat_lv(  
  mapping = NULL,  
  data = NULL,  
  geom = "lv",  
  position = "dodge",
```

```

na.rm = TRUE,
conf = 0.95,
percent = NULL,
k = NULL,
show.legend = NA,
inherit.aes = TRUE,
...
)

StatLv

```

## Arguments

mapping	Set of aesthetic mappings created by <code>aes()</code> . If specified and <code>inherit.aes = TRUE</code> (the default), it is combined with the default mapping at the top level of the plot. You must supply mapping if there is no plot mapping.
data	<p>The data to be displayed in this layer. There are three options:</p> <p>If <code>NULL</code>, the default, the data is inherited from the plot data as specified in the call to <code>ggplot()</code>.</p> <p>A <code>data.frame</code>, or other object, will override the plot data. All objects will be fortified to produce a data frame. See <code>fortify()</code> for which variables will be created.</p> <p>A function will be called with a single argument, the plot data. The return value must be a <code>data.frame</code>, and will be used as the layer data. A function can be created from a formula (e.g. <code>~ head(.x, 10)</code>).</p>
position	Position adjustment, either as a string naming the adjustment (e.g. "jitter" to use <code>position_jitter</code> ), or the result of a call to a position adjustment function. Use the latter if you need to change the settings of the adjustment.
outlier.colour	Override aesthetics used for the outliers. Defaults come from <code>geom_point()</code> .
outlier.shape	Override aesthetics used for the outliers. Defaults come from <code>geom_point()</code> .
outlier.size	Override aesthetics used for the outliers. Defaults come from <code>geom_point()</code> .
outlier.stroke	Override aesthetics used for the outliers. Defaults come from <code>geom_point()</code> .
na.rm	If <code>FALSE</code> (the default), removes missing values with a warning. If <code>TRUE</code> silently removes missing values.
varwidth	if <code>FALSE</code> (default) draw boxes that are the same size for each group. If <code>TRUE</code> , boxes are drawn with widths proportional to the square-roots of the number of observations in the groups (possibly weighted, using the <code>weight</code> aesthetic).
width.method	character, one of 'linear' (default), 'area', or 'height'. This parameter determines whether the width of the box for letter value LV(i) should be proportional to i (linear), proportional to $2^i$ (height), or whether the area of the box should be proportional to $2^i$ (area).
show.legend	logical. Should this layer be included in the legends? <code>NA</code> , the default, includes if any aesthetics are mapped. <code>FALSE</code> never includes, and <code>TRUE</code> always includes. It can also be a named logical vector to finely select the aesthetics to display.

<code>inherit.aes</code>	If FALSE, overrides the default aesthetics, rather than combining with them. This is most useful for helper functions that define both data and aesthetics and shouldn't inherit behaviour from the default plot specification, e.g. <code>borders()</code> .
<code>...</code>	Other arguments passed on to <code>layer()</code> . These are often aesthetics, used to set an aesthetic to a fixed value, like <code>colour = "red"</code> or <code>size = 3</code> . They may also be parameters to the paired geom/stat.
<code>geom, stat</code>	Use to override the default connection between <code>geom_lv</code> and <code>stat_lv</code> .
<code>conf</code>	confidence level
<code>percent</code>	numeric value: percent of data in outliers
<code>k</code>	number of letter values shown

### Format

An object of class `GeomLv` (inherits from `Geom`, `ggproto`, `gg`) of length 6.

An object of class `StatLv` (inherits from `Stat`, `ggproto`, `gg`) of length 5.

### Computed/reported variables

**k** Number of Letter Values used for the display

**LV** Name of the Letter Value

**width** width of the interquartile box

### References

McGill, R., Tukey, J. W. and Larsen, W. A. (1978) Variations of box plots. *The American Statistician* 32, 12-16.

### See Also

[stat\\_quantile](#) to view quantiles conditioned on a continuous variable.

### Examples

```
library(ggplot2)
p <- ggplot(mpg, aes(class, hwy))
p + geom_lv(aes(fill = after_stat(LV))) + scale_fill_brewer()
p + geom_lv() + geom_jitter(width = 0.2)
p + geom_lv(aes(fill = after_stat(LV))) + scale_fill_lv()

# Outliers
p + geom_lv(varwidth = TRUE, aes(fill = after_stat(LV))) + scale_fill_lv()
p + geom_lv(fill = "grey80", colour = "black")
p + geom_lv(outlier.colour = "red", outlier.shape = 1)

# Plots are automatically dodged when any aesthetic is a factor
p + geom_lv(aes(fill = drv))

# varwidth adjusts the width of the boxes according to the number of observations
```

```

ggplot(ontime, aes(UniqueCarrier, TaxiIn + TaxiOut)) +
  geom_lv(aes(fill = after_stat(LV)), varwidth=TRUE) +
  scale_fill_lv() +
  scale_y_sqrt() +
  theme_bw()

ontime$DayOfWeek <- as.POSIXlt(ontime$FlightDate)$wday
ggplot(ontime, aes(factor(DayOfWeek), TaxiIn + TaxiOut)) +
  geom_lv(aes(fill = after_stat(LV))) +
  scale_fill_lv() +
  scale_y_sqrt() +
  theme_bw()

```

---

LVboxplot

*Side-by-side LV boxplots with base graphics*


---

### Description

An extension of standard boxplots which draws k letter statistics. Conventional boxplots (Tukey 1977) are useful displays for conveying rough information about the central 50% of the data and the extent of the data.

### Usage

```

LVboxplot(x, ...)

## S3 method for class 'formula'
LVboxplot(
  formula,
  alpha = 0.95,
  k = NULL,
  perc = NULL,
  horizontal = TRUE,
  xlab = NULL,
  ylab = NULL,
  col = "grey30",
  bg = "grey90",
  width = 0.9,
  width.method = "linear",
  median.col = "grey10",
  ...
)

## S3 method for class 'numeric'
LVboxplot(
  x,
  alpha = 0.95,
  k = NULL,

```

```

perc = NULL,
horizontal = TRUE,
xlab = NULL,
ylab = NULL,
col = "grey30",
bg = "grey90",
width = 0.9,
width.method = "linear",
median.col = "grey10",
...
)

```

### Arguments

x	numeric vector of data
...	passed onto <a href="#">plot</a>
formula	a plotting formula of the form $y \sim x$ , where $x$ is a string or factor. The values of $y$ will be split into groups according to their values on $x$ and separate letter value box plots of $y$ are drawn side by side in the same display.
alpha	if supplied, depth $k$ is calculated such that $(1-\alpha)100$ intervals of an LV statistic do not extend into neighboring LV statistics.
k	number of letter value statistics used
perc	if supplied, depth $k$ is adjusted such that $\text{perc}$ percent outliers are shown
horizontal	display horizontally (TRUE) or vertically (FALSE)
xlab	x axis label
ylab	y axis label
col	vector of colours to use
bg	background colour
width	maximum height/width of box
width.method	one of 'linear', 'height' or 'area'. Methods 'height' and 'area' ensure that these dimension are proportional to the number of observations within each box.
median.col	colour of the line for the median

### Details

For moderate-sized data sets ( $n < 1000$ ), detailed estimates of tail behavior beyond the quartiles may not be trustworthy, so the information provided by boxplots is appropriately somewhat vague beyond the quartiles, and the expected number of “outliers” and “far-out” values for a Gaussian sample of size  $n$  is often less than 10 (Hoaglin, Iglewicz, and Tukey 1986). Large data sets ( $n \approx 10,000 - 100,000$ ) afford more precise estimates of quantiles in the tails beyond the quartiles and also can be expected to present a large number of “outliers” (about  $0.4 + 0.007n$ ).

The letter-value box plot addresses both these shortcomings: it conveys more detailed information in the tails using letter values, only out to the depths where the letter values are reliable estimates of their corresponding quantiles (corresponding to tail areas of roughly  $2^{-i}$ ); “outliers” are defined as a function of the most extreme letter value shown. All aspects shown on the letter-value boxplot are actual observations, thus remaining faithful to the principles that governed Tukey’s original boxplot.

**Examples**

```

n <- 10
oldpar <- par()
par(mfrow=c(4,2), mar=c(3,3,3,3))
for (i in 1:4) {
  x <- rexp(10 ^ (i + 1))
  boxplot(x, col = "grey", horizontal = TRUE)
  title(paste("Exponential, n = ", length(x)))
  LVboxplot(x, col = "grey", xlab = "")
}
par(mfrow=oldpar$mfrow, mar=oldpar$mar)

with(ontime, LVboxplot(sqrt(TaxiIn + TaxiOut) ~ UniqueCarrier, horizontal=FALSE))

```

---

lvtable	<i>Compute table of k letter values for vector x</i>
---------	--

---

**Description**

Compute table of k letter values for vector x

**Usage**

```
lvtable(x, k, alpha = 0.95)
```

**Arguments**

x	input numeric vector
k	number of letter values to compute
alpha	alpha-threshold for confidence level

---

ontime	<i>Ontime Flight Data</i>
--------	---------------------------

---

**Description**

Data set detailing on-time performance of national US flights in January 2015. This data is a subset of the data provided by the US Department of Transportation. The full data as well as archived or more recent data is available for download from [http://www.transtats.bts.gov/DL\\_SelectFields.asp?Table\\_ID=236&DB\\_Short\\_Name=On-Time](http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time).

**Usage**

```
ontime
```



**Format**

A data frame consisting of the variables

**FlightDate** a date variable of the day of the flight

**UniqueCarrier** factor variable of the carrier (using the two letter abbreviation)

**FlightNum** numeric variable of the flight number

**CRSDepTime** scheduled departure time in hhmm format

**DepTime** actual departure time in hhmm format

**CRSArrTime** scheduled arrival time in hhmm format

**ArrTime** actual arrival time in hhmm format

**TaxiOut** numeric variable of the taxi out time in minutes

**TaxiIn** numeric variable of the taxi in time in minutes

**ArrDelay** Arrival delay, in Minutes

**DepDelay** Departure delay, in Minutes

**CarrierDelay** Carrier Delay, in Minutes

**WeatherDelay** Weather Delay, in Minutes

**NASDelay** National Air System Delay, in Minutes

**SecurityDelay** Security Delay, in Minutes

**LateAircraftDelay** Late Aircraft Delay, in Minutes

**References**

[http://www.transtats.bts.gov/DL\\_SelectFields.asp?Table\\_ID=236&DB\\_Short\\_Name=On-Time](http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time)

**Examples**

```
library(ggplot2)
ggplot(ontime, aes(UniqueCarrier, TaxiIn + TaxiOut)) +
  geom_lv(aes(fill = after_stat(LV))) +
  scale_fill_lv() +
  scale_y_sqrt() +
  theme_bw()
```

# Index

## \* datasets

census, 2  
geom\_lv, 3  
ontime, 8

## \* letter-value boxplots

LVboxplot, 6

aes(), 4

borders(), 5

census, 2

determineDepth, 2

fortify(), 4

geom\_lv, 3

GeomLv (geom\_lv), 3

ggplot(), 4

layer(), 5

LVboxplot, 6

lvtable, 8

ontime, 8

plot, 7

scale\_fill\_lv (geom\_lv), 3

stat\_lv (geom\_lv), 3

stat\_quantile, 5

StatLv (geom\_lv), 3