SkillCheck 2.1 geom_histogram(), geom_density(), geom_dotplot()

AJ Anzai, Kim Nguyen, Lorenzo Pessi

diamonds

| carat <dbl>	cut <ord>	color <ord>	clarity <ord>	depth <dbl>	table <dbl>	price <int>	x <dbl>	y <dbl>	z <dbl>
0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75
0.24	Very Good	J	VVS2	62.8	57.0	336	3.94	3.96	2.48
0.24	Very Good	I	VVS1	62.3	57.0	336	3.95	3.98	2.47
0.26	Very Good	H	SI1	61.9	55.0	337	4.07	4.11	2.53
0.22	Fair	E	VS2	65.1	61.0	337	3.87	3.78	2.49
0.23	Very Good	H	VS1	59.4	61.0	338	4.00	4.05	2.39

geom_histogram()

AJ Anzai

- Characteristics of data include numeric and integer inputs


```r
{r}
diamonds
x.axis <- diamonds$cut
geom_histogram(x= x.axis)

ggplot(diamonds, aes(x= depth)) + geom_histogram(binwidth = 5)
ggplot(diamonds, aes(x= depth)) + geom_histogram(color="blue",
fill="orange")
```
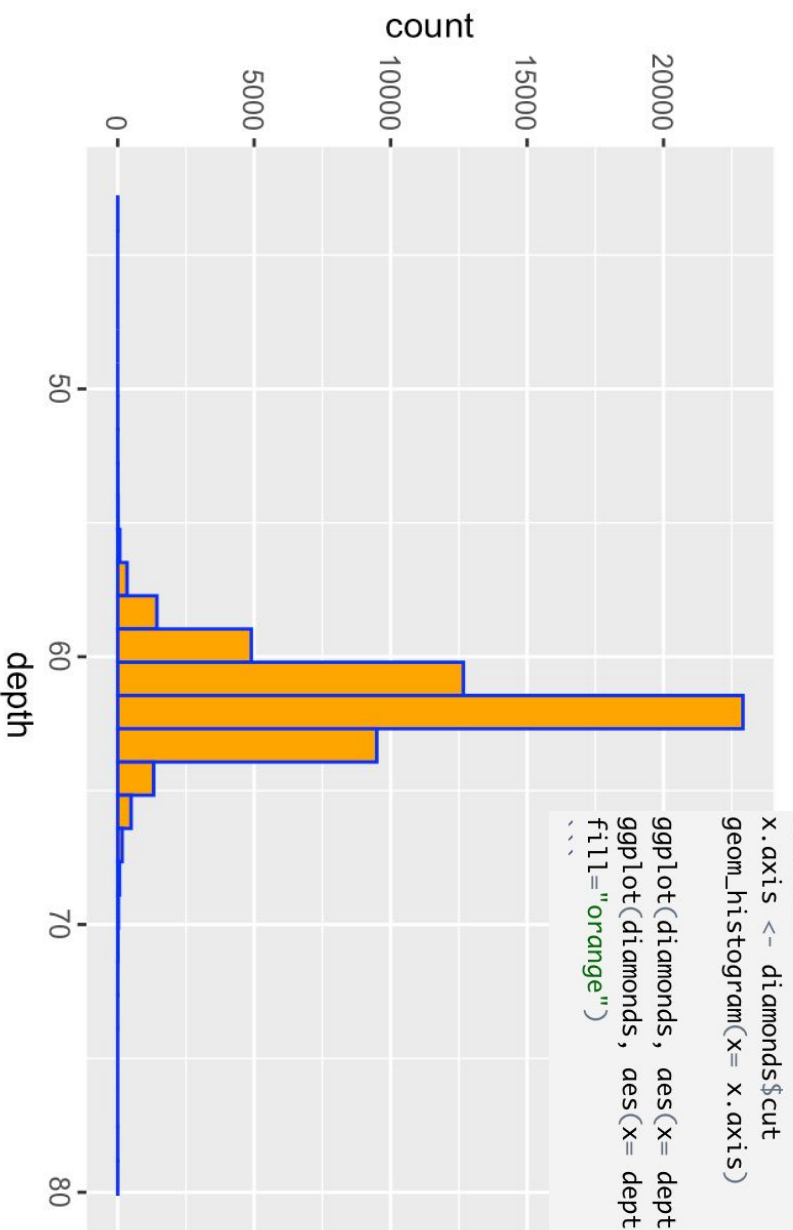
★ Geom_histogram divides the x axis into bins and counts the number of observations in each bin.

geom_histogram()

AJ Anzai

Arguments:

1) Mapping- allows the user to create a histogram map using default settings
2) Data- a data frame will override the plot data
3) Position- position adjustment
4) ...- to set an aesthetic to a fixed value
5) Na.rm- removes missing values from data set with or without a warning
6) Show.legend- will provide a legend
7) Inherit.aes- overrides default aesthetics
8) Binwidth- width of the bins
9) Bins- number of bins
10) Orientation- how the plot is structured
11) Geom, Stat- overrides default connection between geom_histogram() and geom_freqpoly()
12) Center, Boundary- bin position specifiers
13) Breaks- supplies a numeric vector giving the bin boundaries
14) Closed- whether the right or left edge of bin is included in the bin
15) Pad- adds empty bins

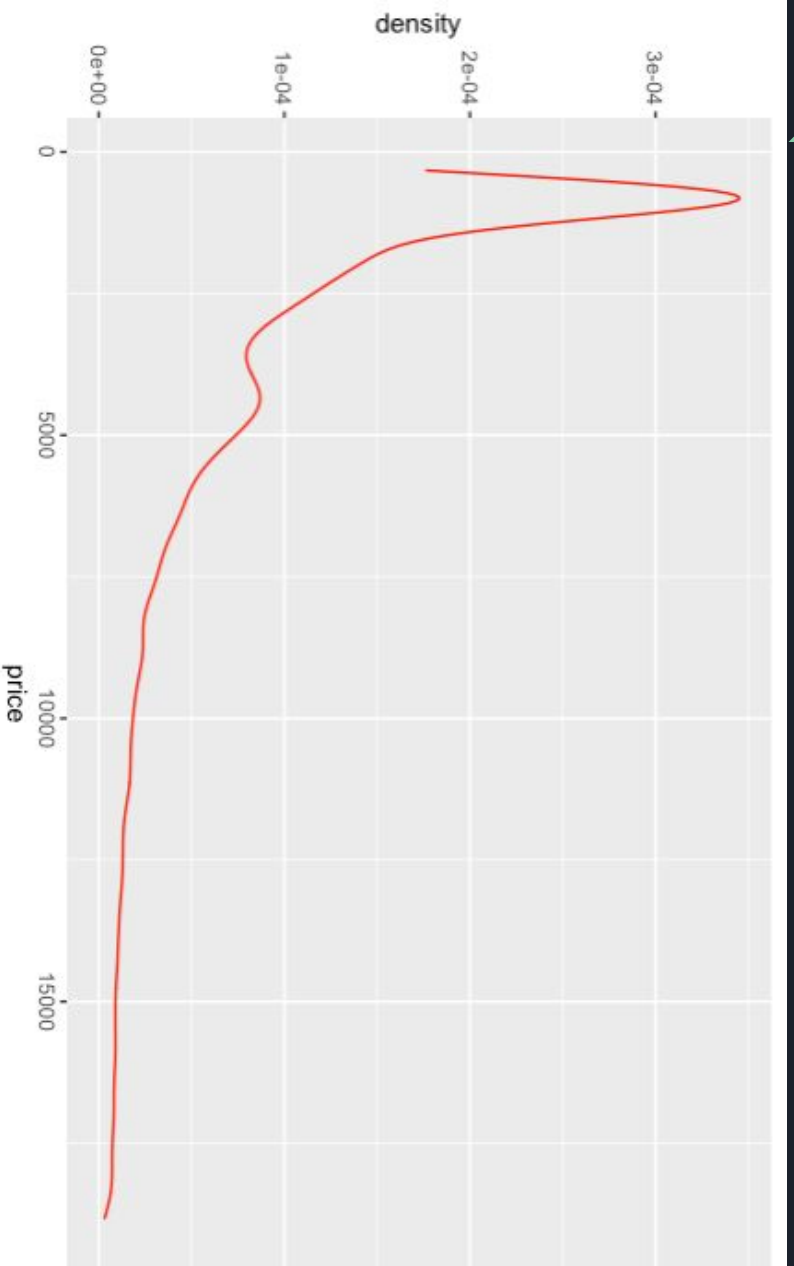# geom_density()

- Characteristics: numeric, integer
  - Cannot accept categorical values!
- Data style: neither wide nor long
  - Only looks at one value at a time

geom_density()

```{r}
diamonds
ggplot(diamonds, aes(x = price)) + geom_density(col = "red")
```


★ geom_density is a geom_histogram drawn with a smooth curve instead of bars (divides the x axis into bins and counts the number of observations in each bin)

geom_density() - Arguments

- mapping - Creates visualization of data
- data - Dataframe overrides plot data
- position - position of the graph
- na.rm - removes missing values
- orientation - orientation of the layer(x or y)
- show.legend - show or hide legend
- inherit.aes - keep or override default aesthetics
- outline.type - outlines the graph
- geom, stat - override the default connection between geom_density() and stat_density().

Geom_dotplot()

```r
{r}
diamonds
ggplot(diamonds, aes(x = depth)) + geom_dotplot(method="histodot", binwidth = 0.01, col = "cyan1")
+ theme_minimal()
```
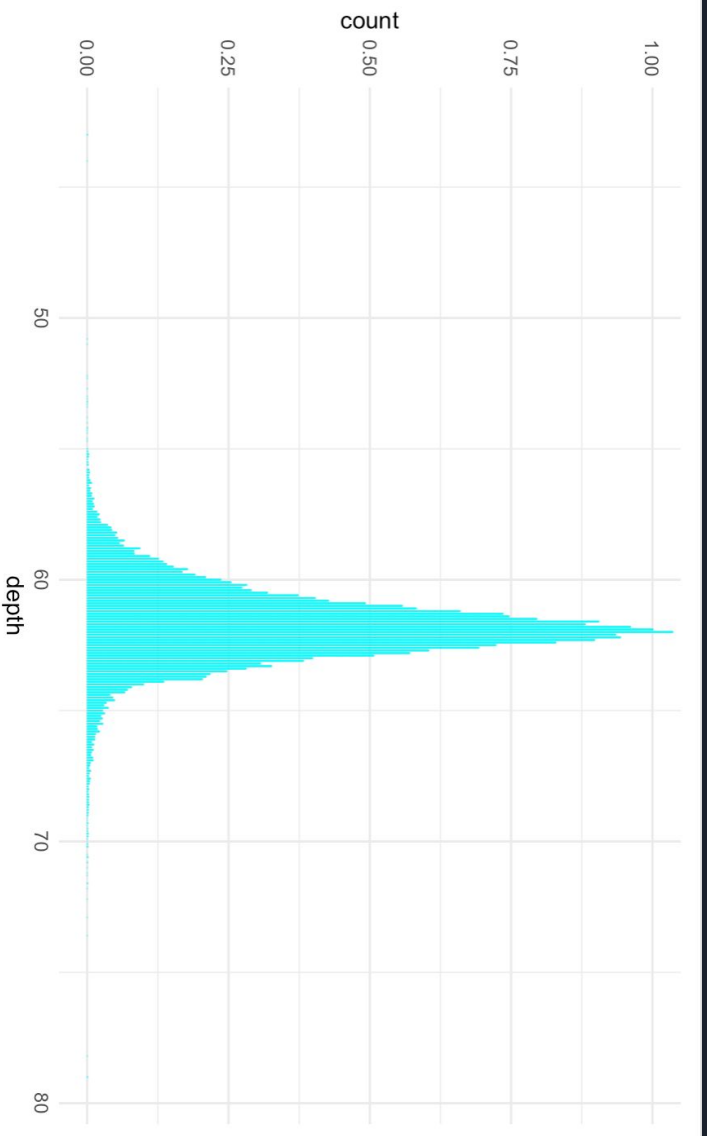
- Data is being cut in sets of intervals known as bins
- The dot width corresponds to the width of a bin (data interval), each dot represents 1 observation and they are stacked
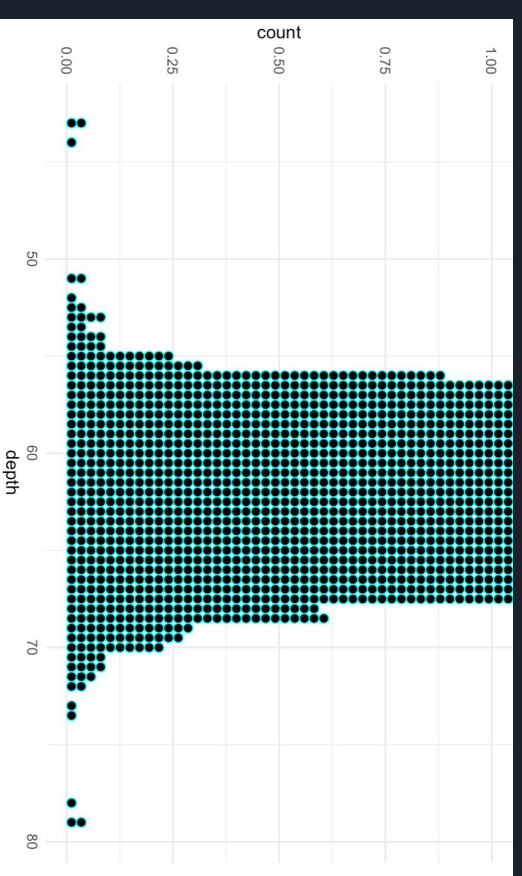
Geom_dotplot() Arguments

- **Mapping:** creates visualization of data through graph
- **Data:** default is NULL, dataframe overrides plot data
- **Position:** adjusts position
- **Binwidth:** Adjusts the bin width, resulting in a change in size of the dots
- **Binaxis:** which axis the bin is set to, default is x and can be changed to y
- **Method:** "dotdensity" is default, can be changed to "histodot" for fixed bin width
- **Binpositions:** determines bin position for "dotdensity" using "bygroup"
- **Stackdir:** direction the dots are stacked, default is "up", change to "down", "center", or "centerwhole"
- **Stackratio:** how close the dots are stacked to each other, default is 1, <1 will overlap dots
- **Dotsize:** dot diameter in relation to binwidth, default is 1
- **Stackgroups:** allows dots to be stacked across groups
- **Origin:** in "histodot", gives origin of first bin
- **Right:** whether intervals should be closed on the right (a,b] or shouldn't [a,b)
- **Width:** in binaxis "y", correlates to spacing of dot stacks
- **Drop:** set to TRUE, results in all bins with zero counts to be eliminated
- **Na.rm:** FALSE -> missing values removed with a warning vs TRUE -> removed silently
- **Show.legend:** logical assessing whether a layer is in the legend, FALSE=never, TRUE=always
- **Inherit.aes:** FALSE=overrides default aesthetics

more information

1. http://www.sthda.com/english/wiki/ggplot2-histogram-plot-quick-start-guide-r-software-and-data-visualization
2. https://ggplot2.tidyverse.org/reference/geom_density.html
3.