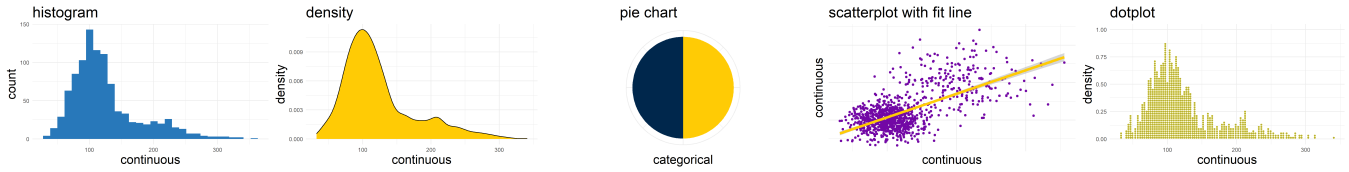
How to Choose a Chart

A Statistically Motivated Guide

Andy Grogan-Kaylor

April 13, 2021



How to Choose a Chart

Choosing the right chart to represent your data can be a daunting process. I believe that a *starting point* for this thinking is some basic statistical thinking about the *type* of variables that you have. At the broadest level, variables may be conceptualized as *categorical* variables, or *continuous* variables.

- *categorical variables* represent unordered categories like *gender*, or *religious affiliation*.
- *continuous variables* represent a continuous scale like a *mental health scale*, or a *measure of neighborhood quality*.

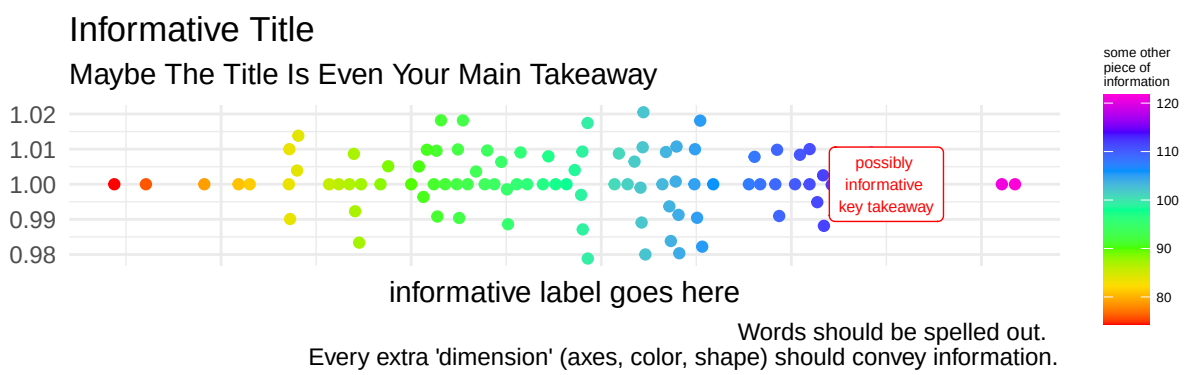
Once we have discerned the type of variable that have, there are two followup questions we may ask before deciding upon a chart strategy:

- Is our graph about **one thing at a time**?
  - How much of *x* is there?
  - What is the distribution of *x*?
- Is our graph about **two things at a time**?
  - What is the relationship of *x* and *y*?
  - How are *x* and *y* associated?

A Few Notes

A Note About Graph Labels

Graphs should have clear titles and labels.



Words should be spelled out.
Every extra 'dimension' (axes, color, shape) should convey information.

A Note About Software

The principles of graphing discussed in this document transcend any particular software package, and could be implemented in many different software packages, such as SPSS, SAS, Stata, or R.

The graphs in these particular examples use ggplot2, a graphing library in R. ggplot2 graph syntax can be formidably complex, with a somewhat steep learning curve. More information about ggplot can be found here.

A Note About Graph Colors

This document uses colors based upon official University of Michigan colors. Using colors that match the design scheme of your organization may be helpful.

A Simulated Data File of Continuous and Categorical Data

A few randomly selected observations…

	x	y	z	u	v	w	s	q
783	82.14	110.9	102.4	Group B	Group A	Group A	Group 2	102.1
496	59.31	115.9	150.3	Group A	Group A	Group A	Group 3	89.31
5	109.2	103.6	116.2	Group A	Group A	Group A	Group 2	129.2
362	201.3	102.8	80.81	Group A	Group B	Group B	Group 2	221.3
478	94.69	211.3	71.1	Group A	Group A	Group B	Group 1	104.7
840	190.4	262	97.41	Group B	Group B	Group B	Group 3	220.4
842	71.73	99.34	122.7	Group B	Group B	Group A	Group 2	91.73
740	94.38	150	88.92	Group B	Group A	Group A	Group 2	114.4

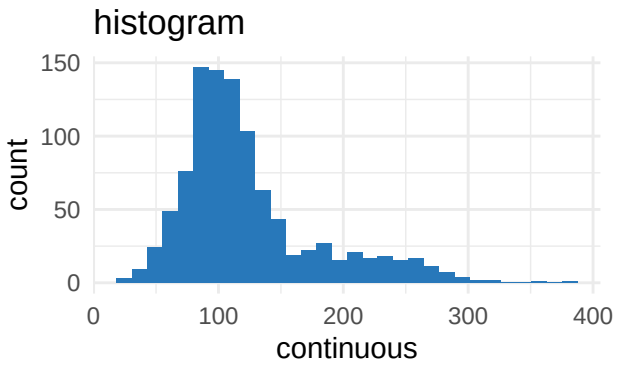
	x	y	z	u	v	w	s	q
537	96.01	53.32	100.9	Group A	Group A	Group A	Group 1	106
664	83.83	99.58	98.59	Group A	Group A	Group A	Group 3	113.8

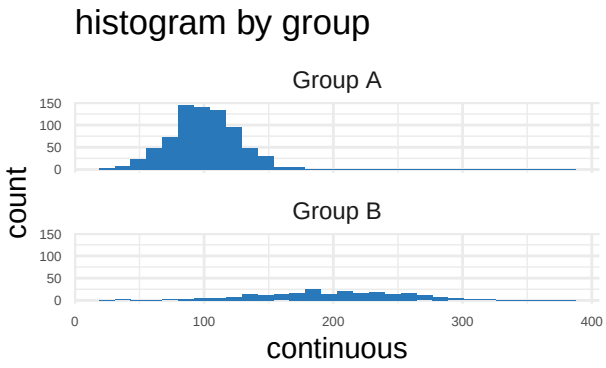
One Thing At A Time

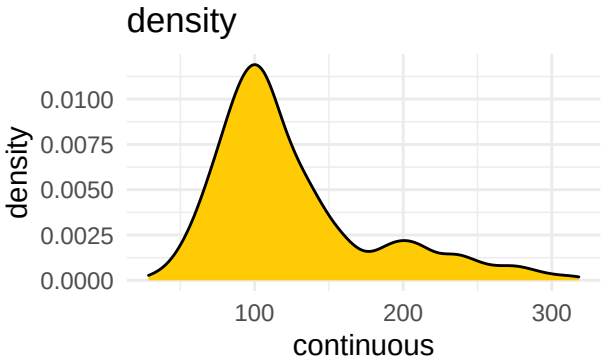
Two Things At A Time

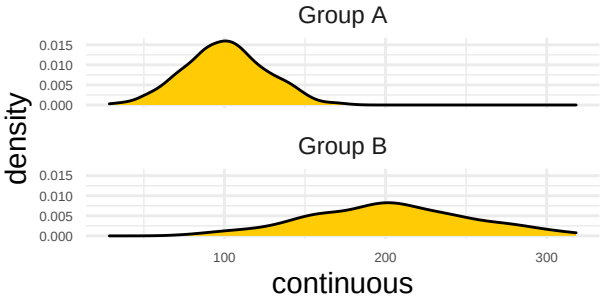
Continuous

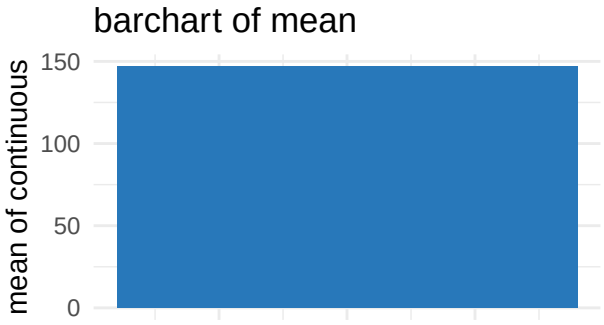
Continuous By Categorical

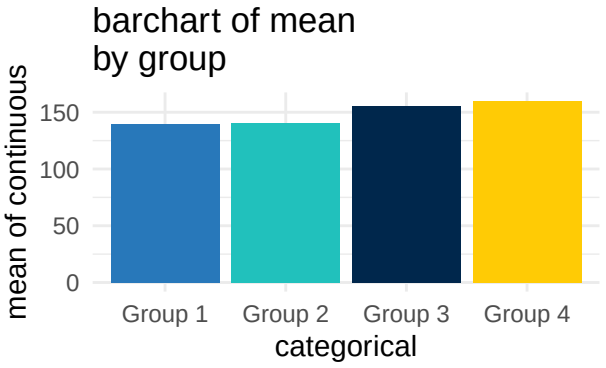




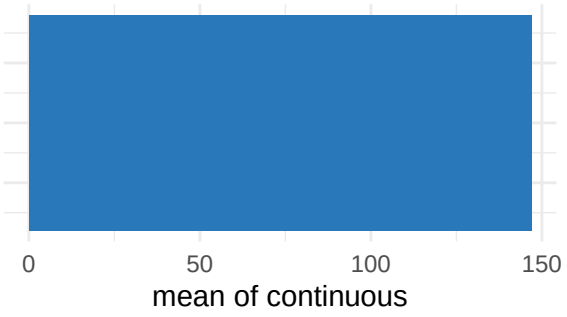




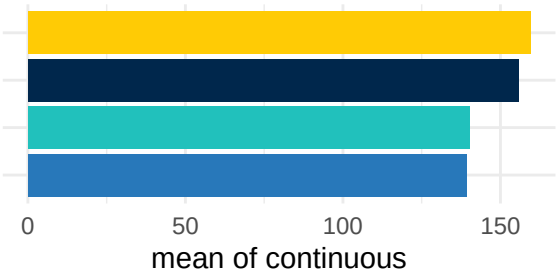




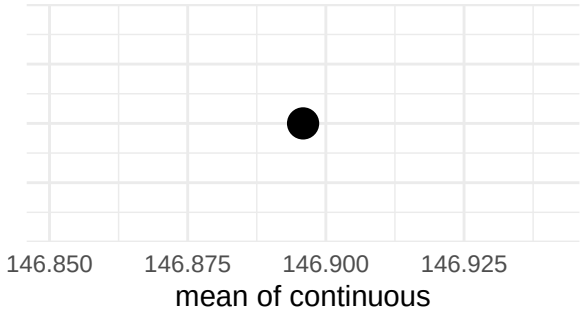
horizontal barchart of mean



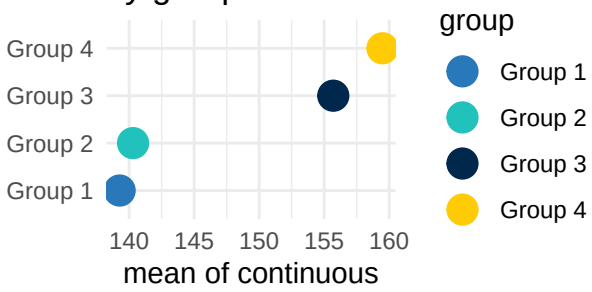
horizontal barchart of mean by group



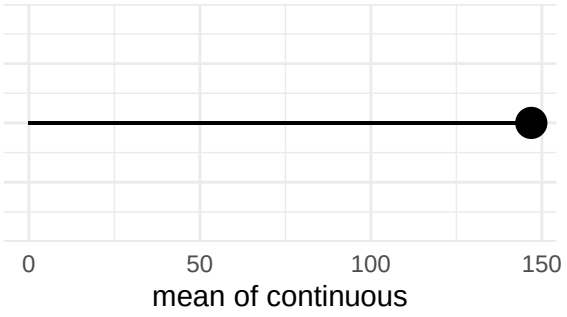
horizontal dotchart of mean



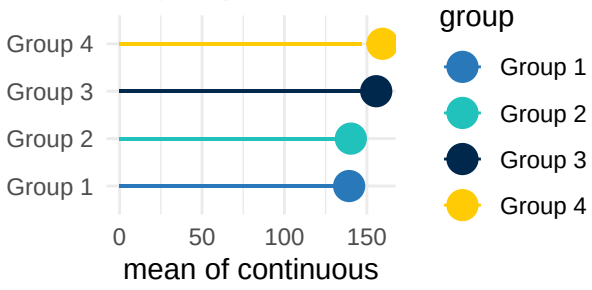
horizontal dotchart of mean by group



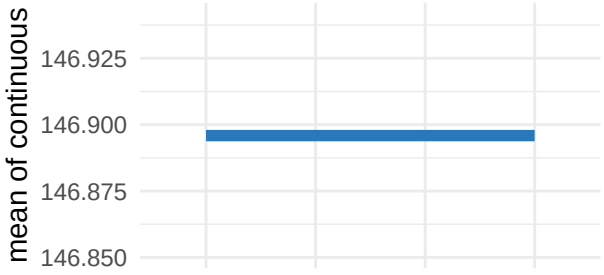
horizontal lollipop chart of mean



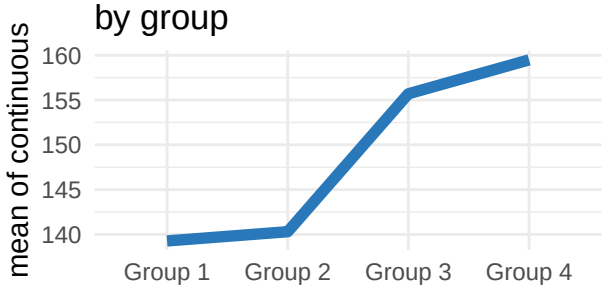
horizontal lollipop chart of mean by group



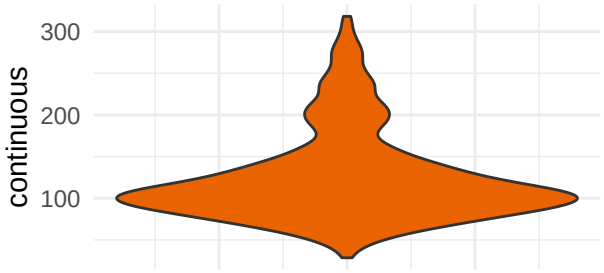
linechart of mean



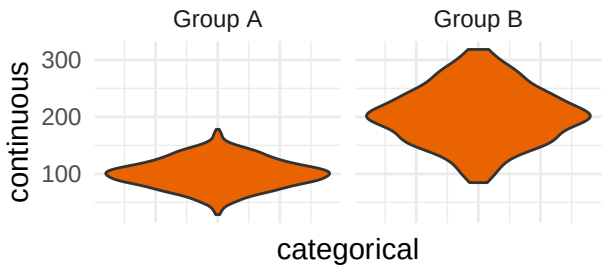
linechart of mean by group



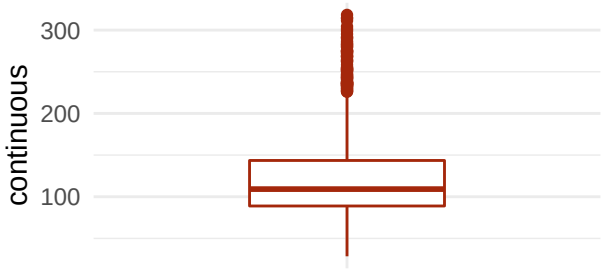
violin plot



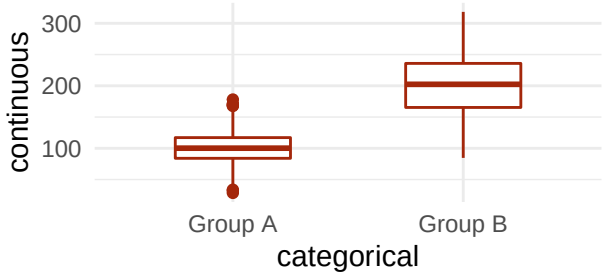
violin plot
by group



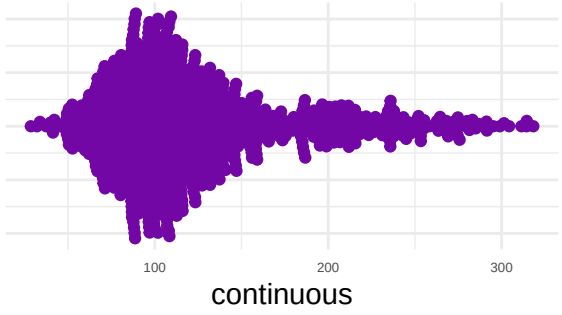
boxplot



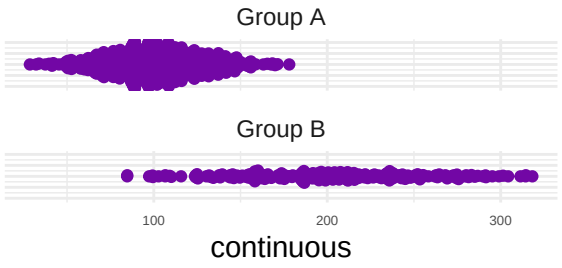
boxplot
by group



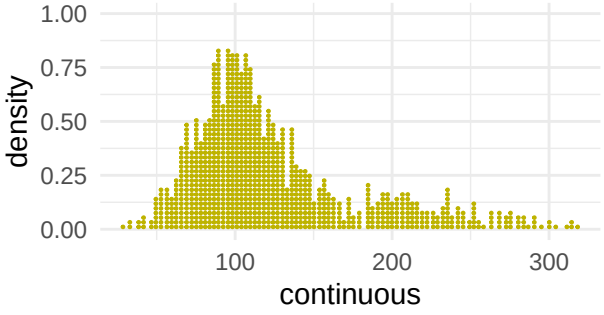
beeswarm plot



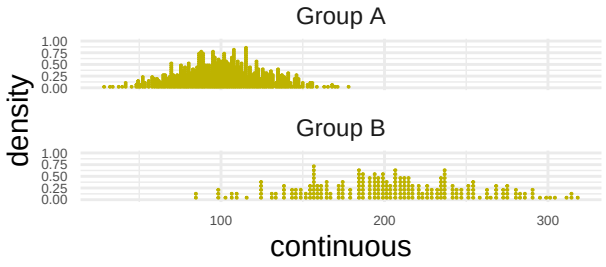
beeswarm plot
by group



dotplot



dotplot
by group



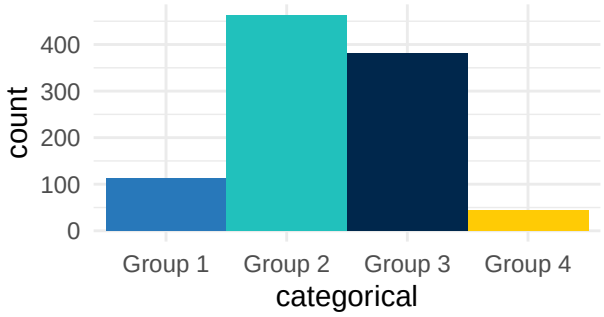
One Thing At A Time

Two Things At A Time

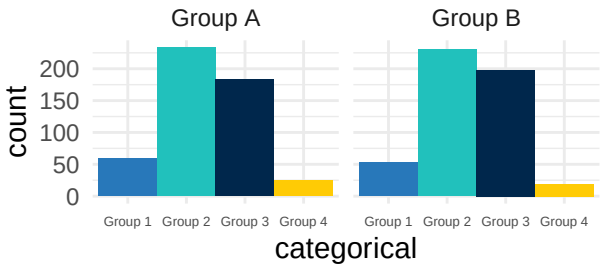
Categorical

Categorical By Categorical

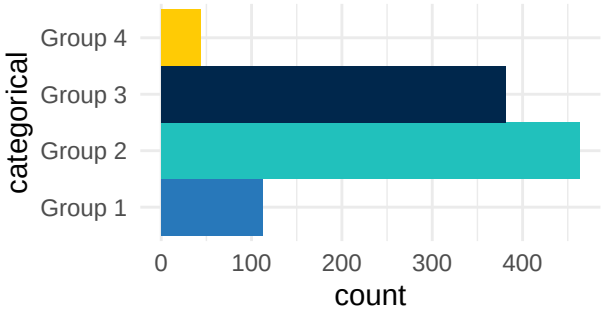
bar chart



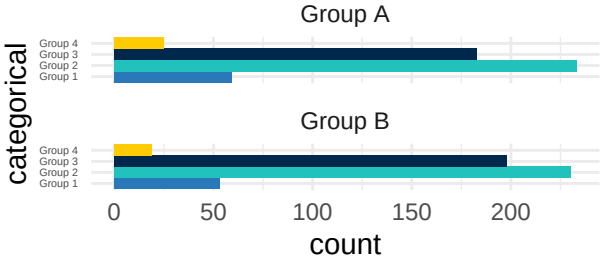
bar chart
by group



horizontal bar chart

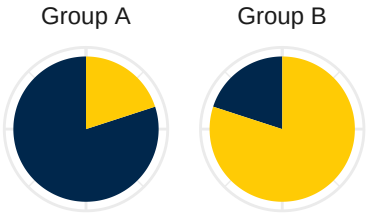


horizontal bar chart
by group



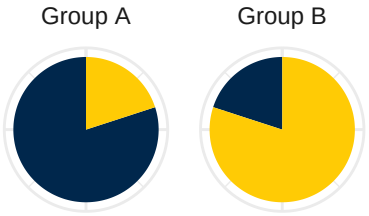
pie chart

pie chart
by group

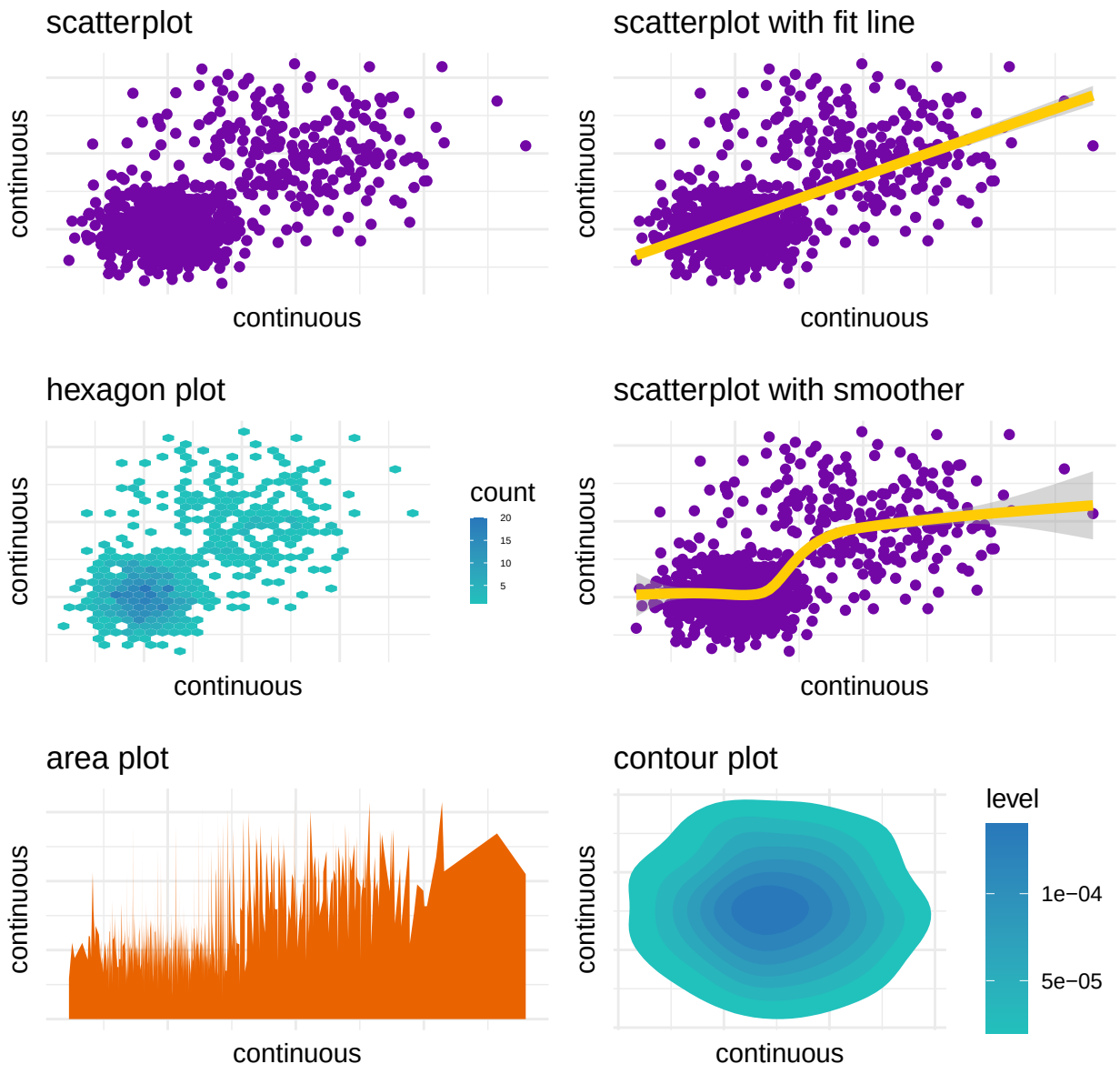


doughnut chart

doughnut chart
by group



Continuous by Continuous



Graphics made with the ggplot2 graphing library created by Hadley Wickham.

Available online at https://agrogan1.github.io/

Last updated: `April 13 2021` at `14:42`