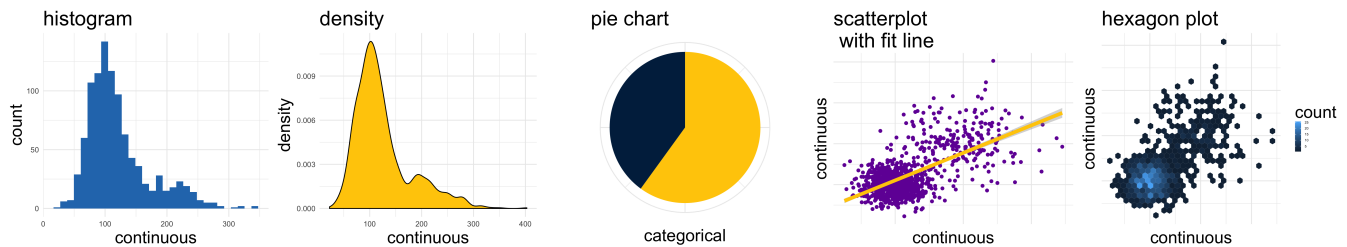


How to Choose a Chart

A Statistically Motivated Guide

Andy Grogan-Kaylor

June 19, 2017



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*.

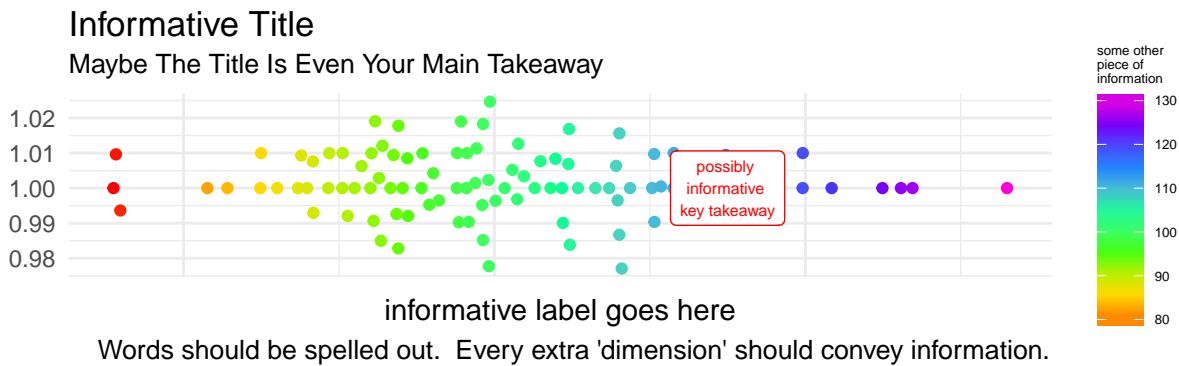
Once we have discerned the type of variable that we have, there are two followup questions we may ask before deciding upon a chart strategy:

- 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.



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.

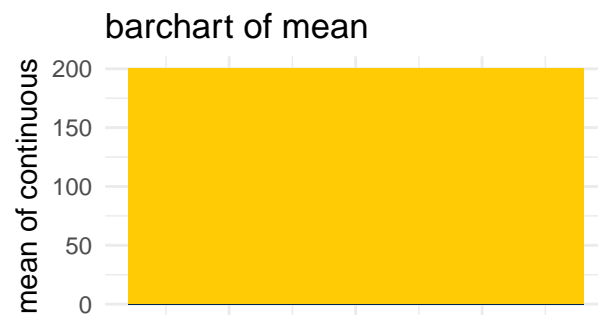
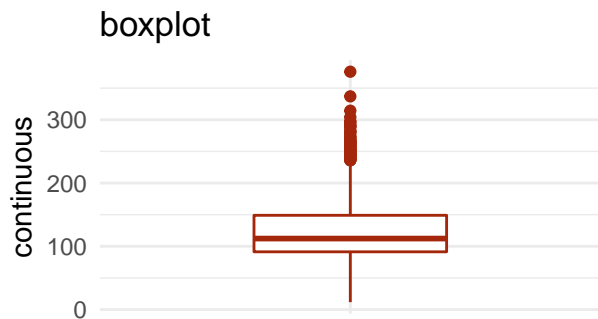
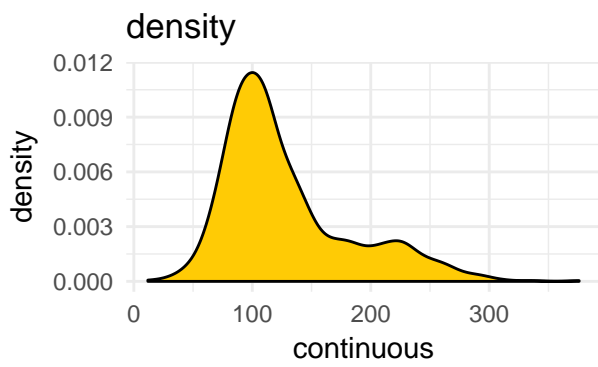
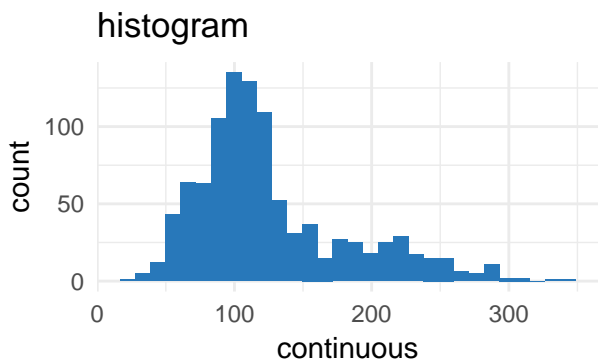
A Simulated Data File of Continuous and Categorical Data

The first few observations...

x	y	z	u	v	w
131.2	88.41	74.47	Group A	Group A	Group A
129.9	99.74	82.12	Group A	Group A	Group A
59.86	139.3	113.9	Group A	Group A	Group A
64.87	132.4	80.68	Group A	Group A	Group A
99.07	96.5	149.1	Group A	Group A	Group A
104.9	83.82	114.6	Group A	Group A	Group A

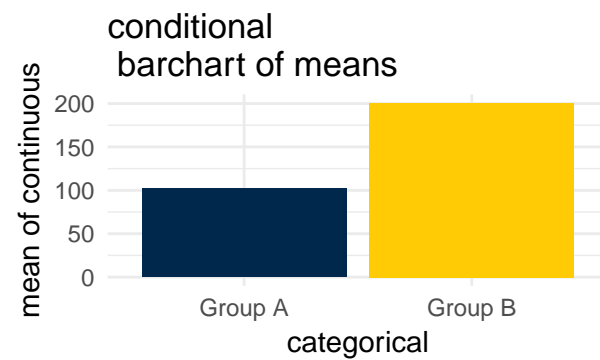
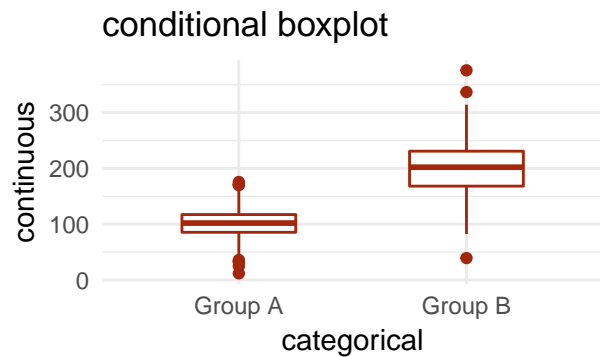
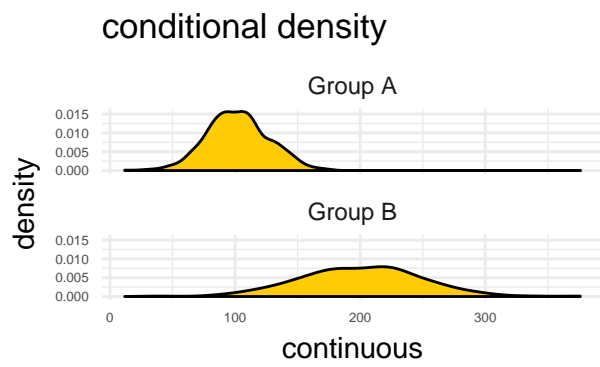
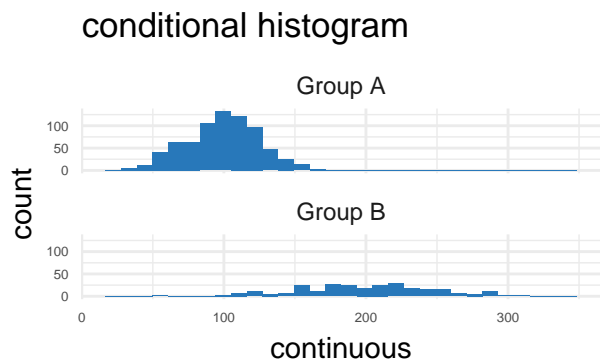
One Thing At A Time

Continuous

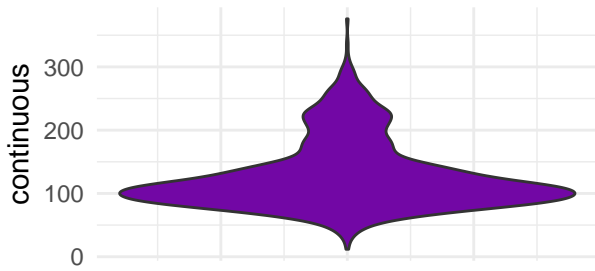


Two Things At A Time

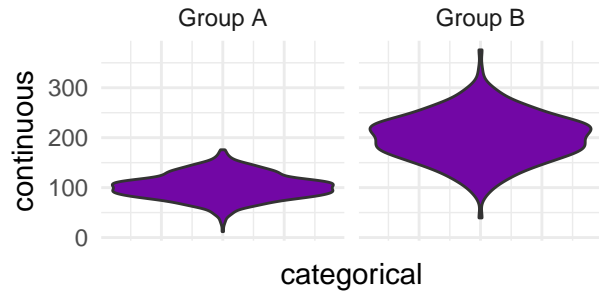
Continuous By Categorical



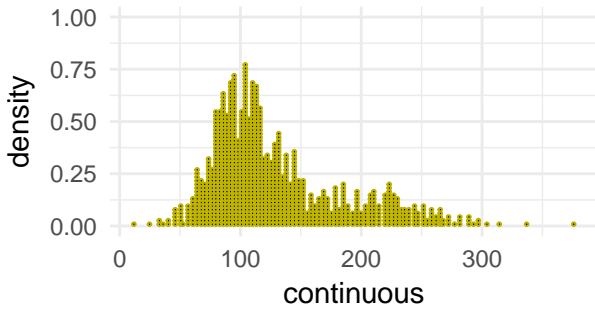
violin plot



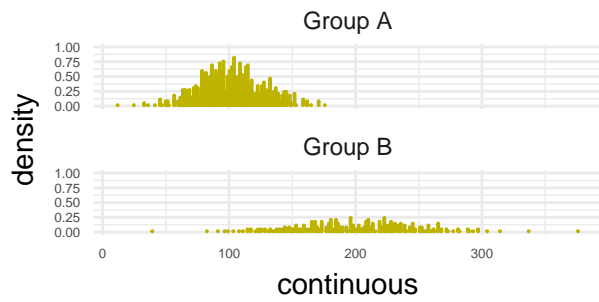
conditional violin plot



dotplot



conditional dotplot



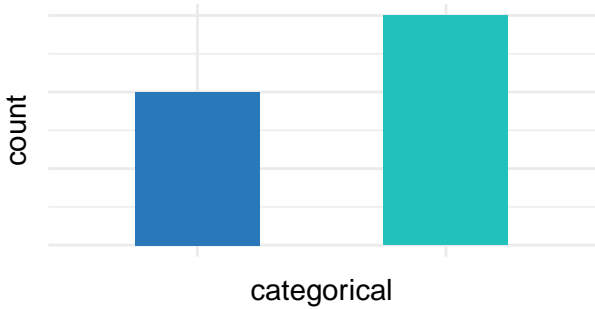
One Thing At A Time

Two Things At A Time

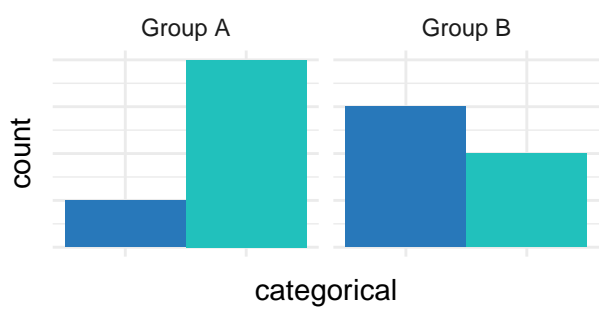
Categorical

Categorical By Categorical

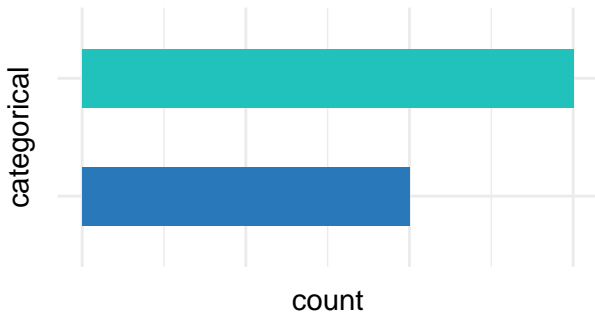
bar chart



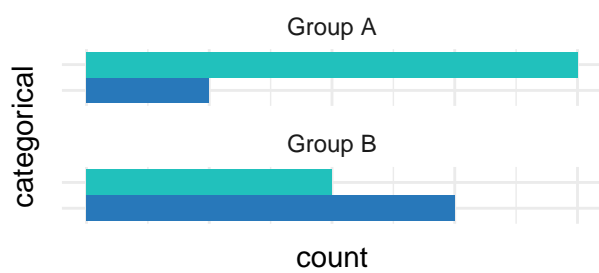
conditional bar chart



horizontal bar chart



conditional horizontal bar chart



pie chart

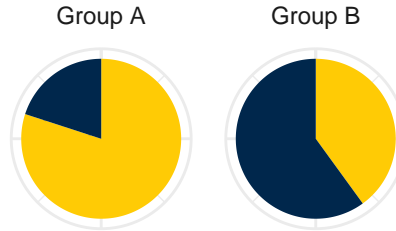


categorical
doughnut chart

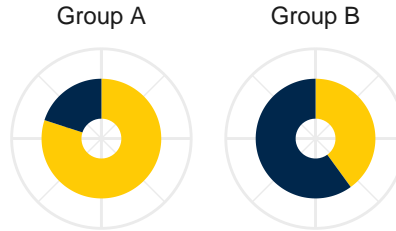


categorical

conditional pie chart



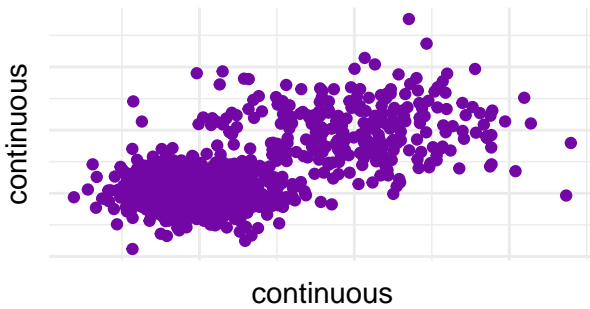
categorical
conditional doughnut chart



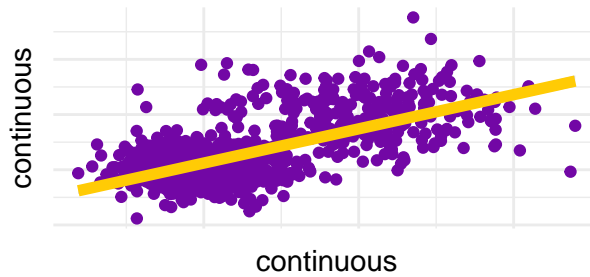
categorical

Continuous by Continuous

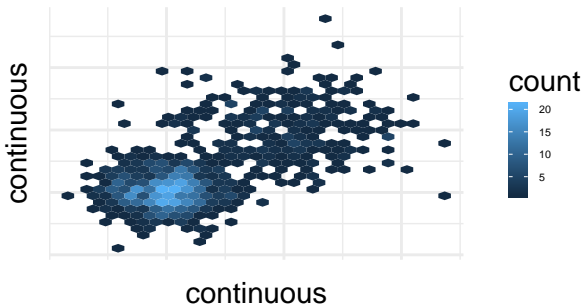
scatterplot



scatterplot
with fit line



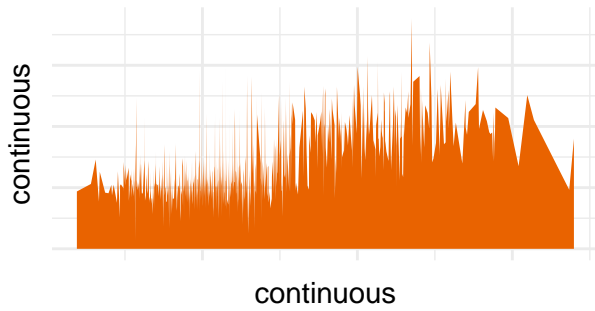
hexagon plot



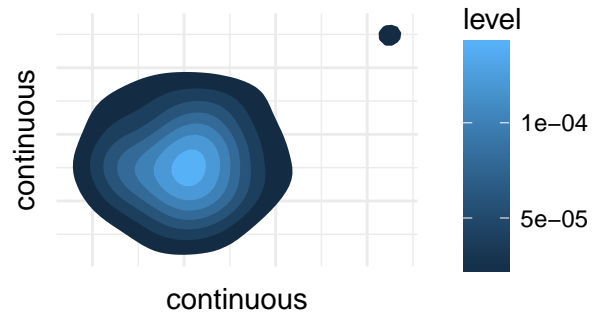
smoother



area plot



contour plot



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

Available online at <https://agroganweb.wordpress.com/data-visualization-dataviz/>



How to Choose a Chart by Andrew Grogan-Kaylor is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. You are welcome to download and use this handout in your own classes, or work, as long as the handout remains properly attributed.

Last updated: June 19 2017 at 12:04