



Day 13 of #100daysofmathandstats: Exploration of two or more variables

By Harsh Kathiriya



Outline

- Contingency table
- Hexagonal binning
- Contour plot
- Violin plot

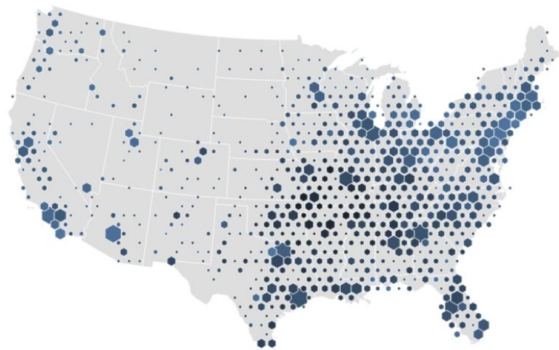
Contingency table

- A tally of counts between **two or more categorical variables**.
- A contingency table provides a way of **portraying data** that can facilitate calculating probabilities.
- A random set of 100 people who have pets were polled to see if there was a correlation between gender and whether they had a dog or a cat. This is a contingency table outlining the data.

	Dog	Cat	Total
Male	42	10	52
Female	9	39	48
Total	51	49	100

Hexagonal binning

- A plot of two numeric **variables with the records** binned into hexagons.
- Hexagonal Binning is another way to manage the problem of having too many points that start to overlap.
- The most evident is that **hexagons** are more similar to circle than square.
This translates in more efficient data aggregation around the bin center.





Real examples of correlation coefficient

- Positive correlations

- The more money you save, the more financially secure you feel.
- As the temperature goes up, ice cream sales also go up.
- The more gasoline you put in your car, the farther it can go.

- Negative correlations

- Time Spent Watching TV vs. Exam Scores
- Time Spent Running vs. Body Fat
- Less study time vs chance of getting high scores

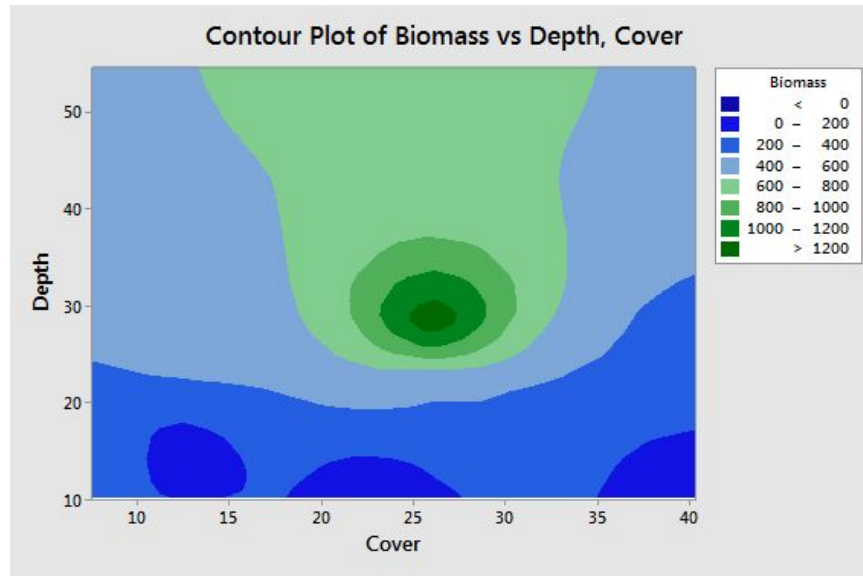


Contour plot

- A plot showing the density of two numeric variables like a **topographical** map.
- Use contour plots to display the relationship between two independent variables and a dependent variable.
- Contour plots are particularly helpful when you need to identify combinations of X and Y that produce beneficial or required values of Z

Example of contour plot

Biologist studies the effect of stream depth and canopy cover on fish biomass.

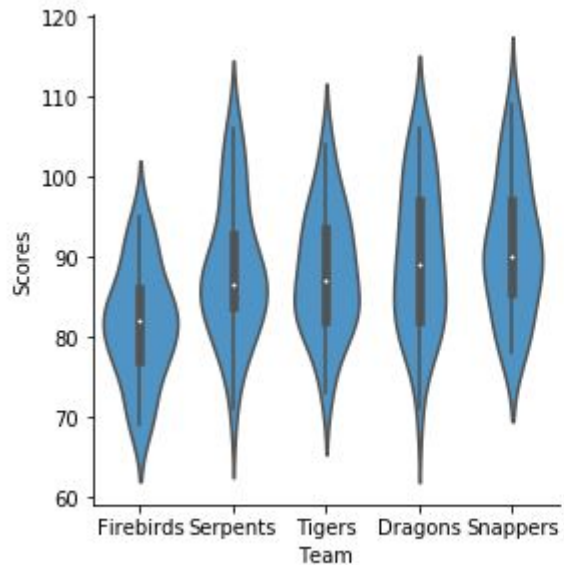




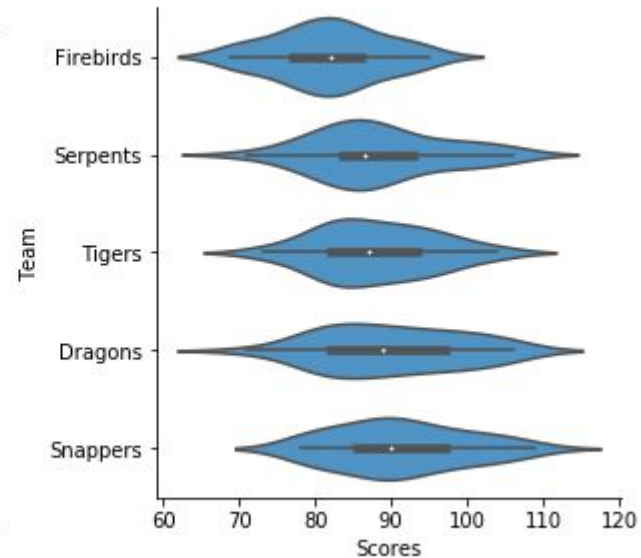
Violin plot

- **Similar to a boxplot** but showing the density estimate
- A violin plot depicts distributions of numeric data for one or more groups using density curves.
- Violin plots are used when you want to **observe the distribution** of numeric data, and are especially useful when you want to make a comparison of distributions **between multiple groups**.

Example of violin plot



Vertical violin plot



Horizontal violin plot



Thank you

Github Link: <https://github.com/harsh9898/100daysofstatandmath>

Don't forget to post your queries or feedbacks on the post.

Share or like for the benefit of others.