

# Summarizing Numerical Data Visualization



DASC 512

# Overview

- Intro to visualization in Python
- Rug Plot
- Histogram
- Kernel Density Estimation (KDE) Plot
- Box Plot
- Scatterplot
- Combined plots
- Finalizing your plots

# Quantitative or Numerical Data

Quantitative data is numerical and ordinal, interval, or ratio

Ordinal – Order has meaning but addition/subtraction does not

- Example: Ranked preference in polls

Interval – Addition/subtraction has meaning but division does not

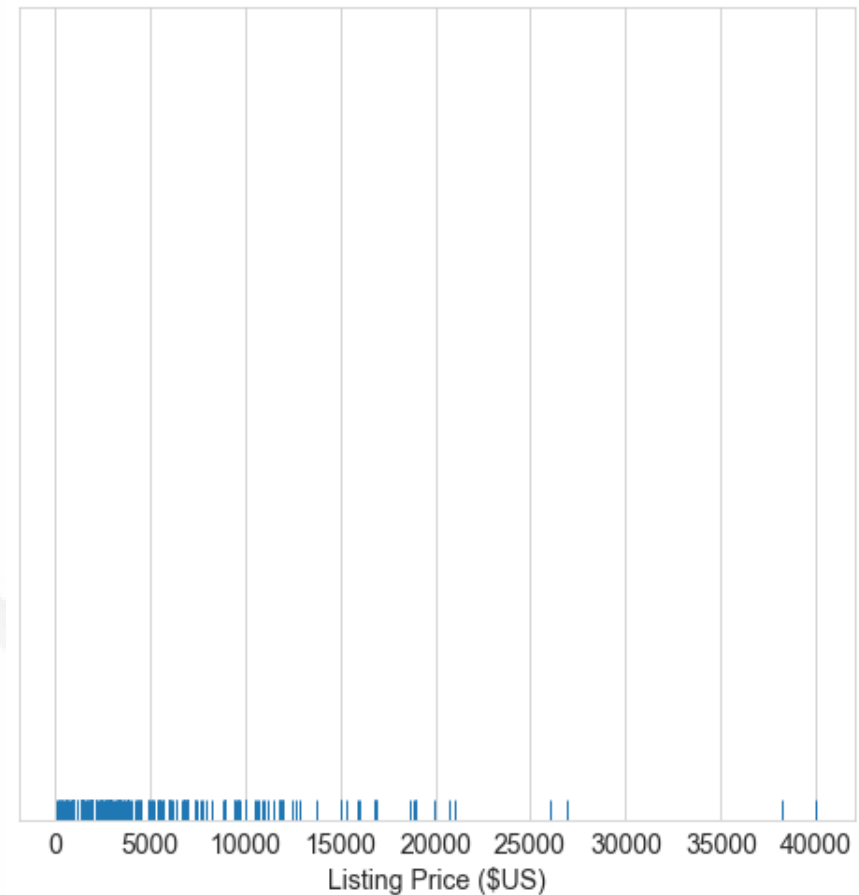
- Example: Temperature in Celsius or Fahrenheit

Ratio – All mathematical operators have meaning

- Example: Height, weight, Kelvin, length, etc.

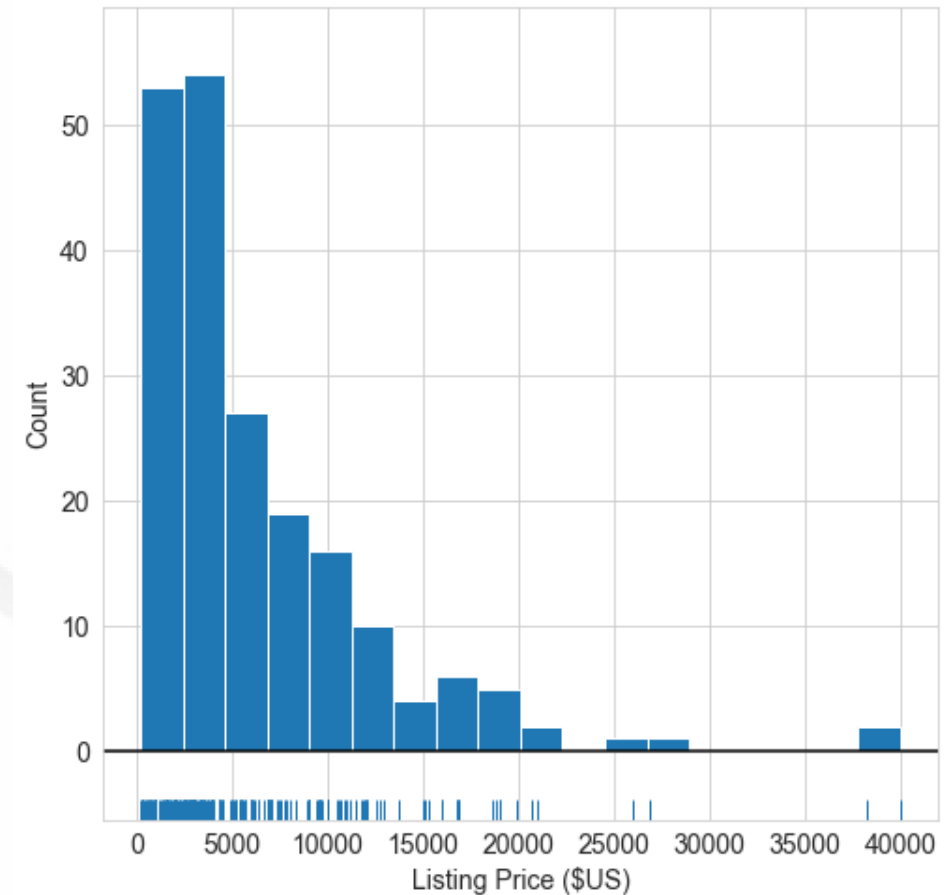
# Rug Plot

- Similar to a dot plot
- Every hash on the x-axis is a single data point
- Not particularly useful on its own



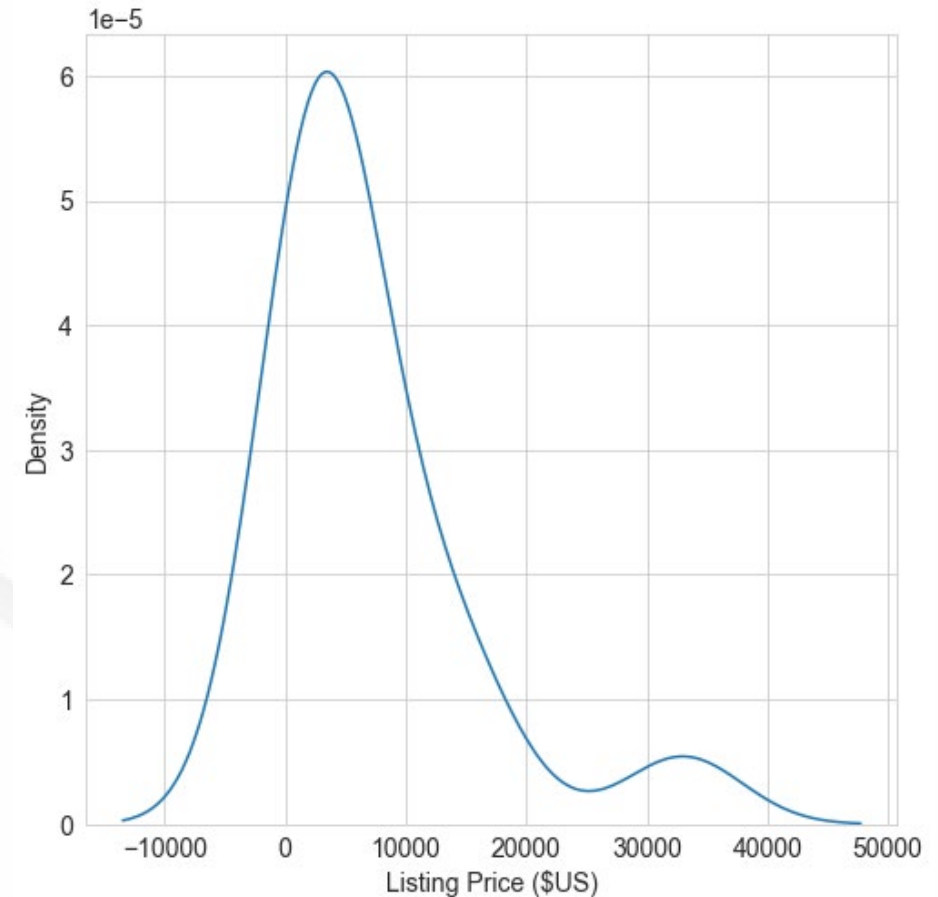
# Histogram

- Data binned into intervals
- Height indicates number of cases with values in each bin
- Number/width of bins depends on data and what is being highlighted



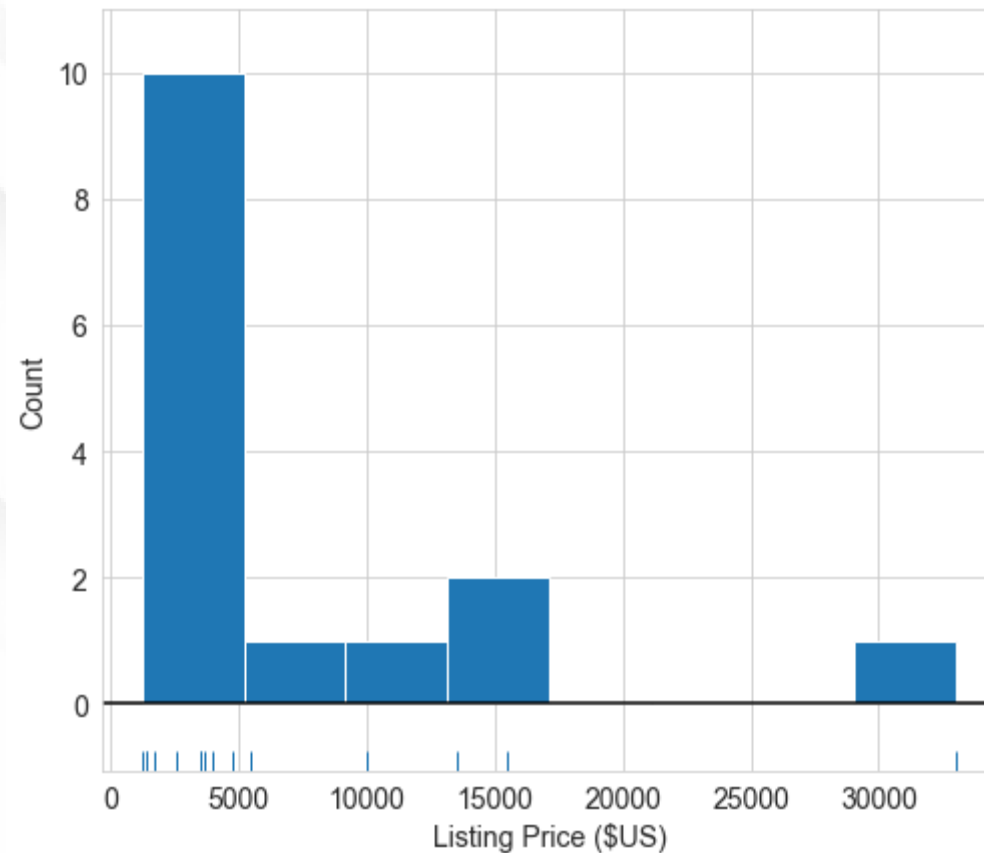
# Kernel Density Estimation

- Smoothed version of a histogram
- May be misleading if used without care and understanding



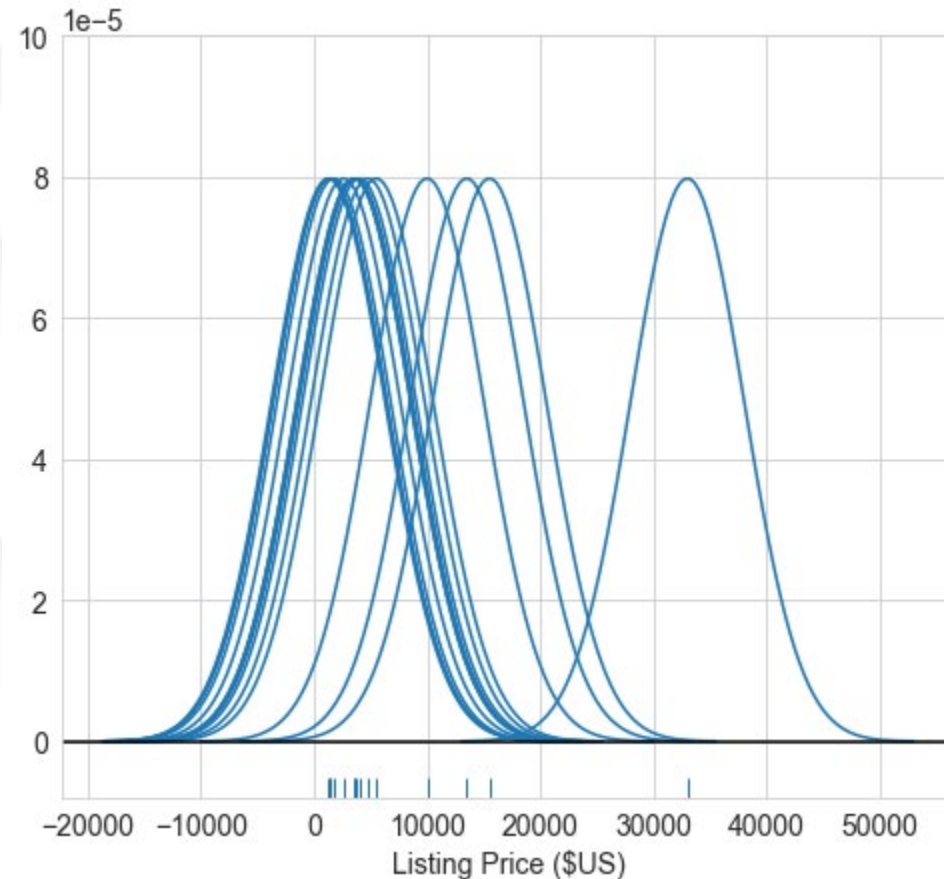
# Kernel Density Estimation

- Start with a rug plot



# Kernel Density Estimation

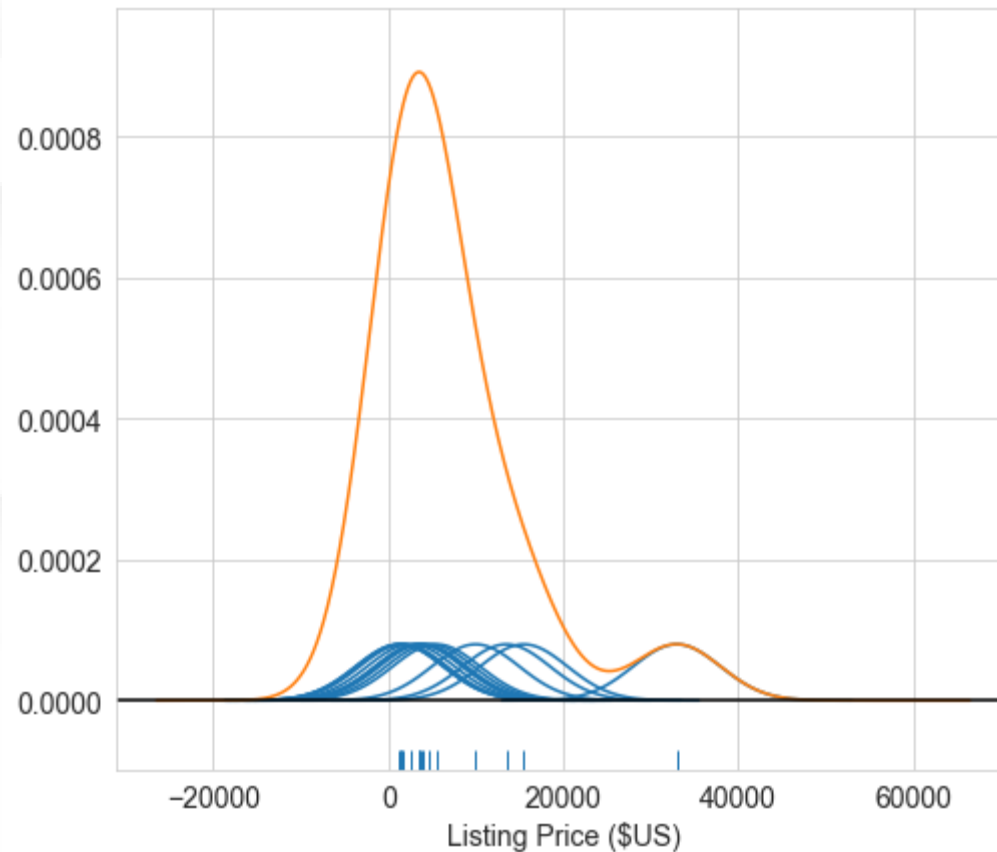
- Add a bell curve centered at each hash on the rug plot



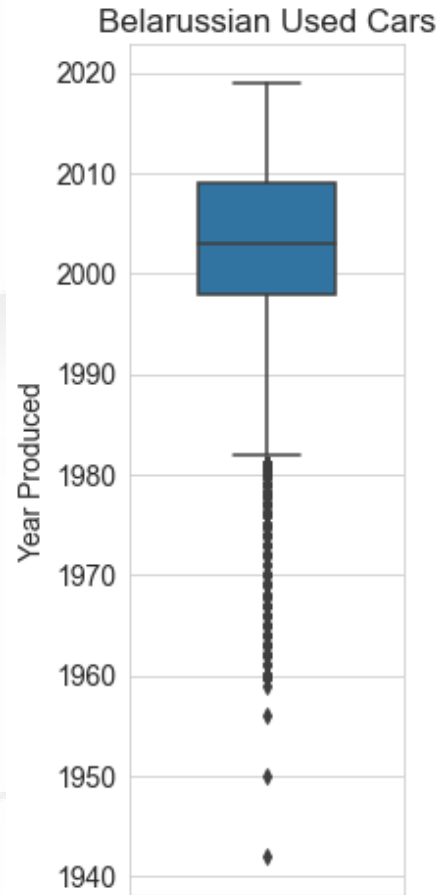


# Kernel Density Estimation

- Now add all the bell curves together, and voila!



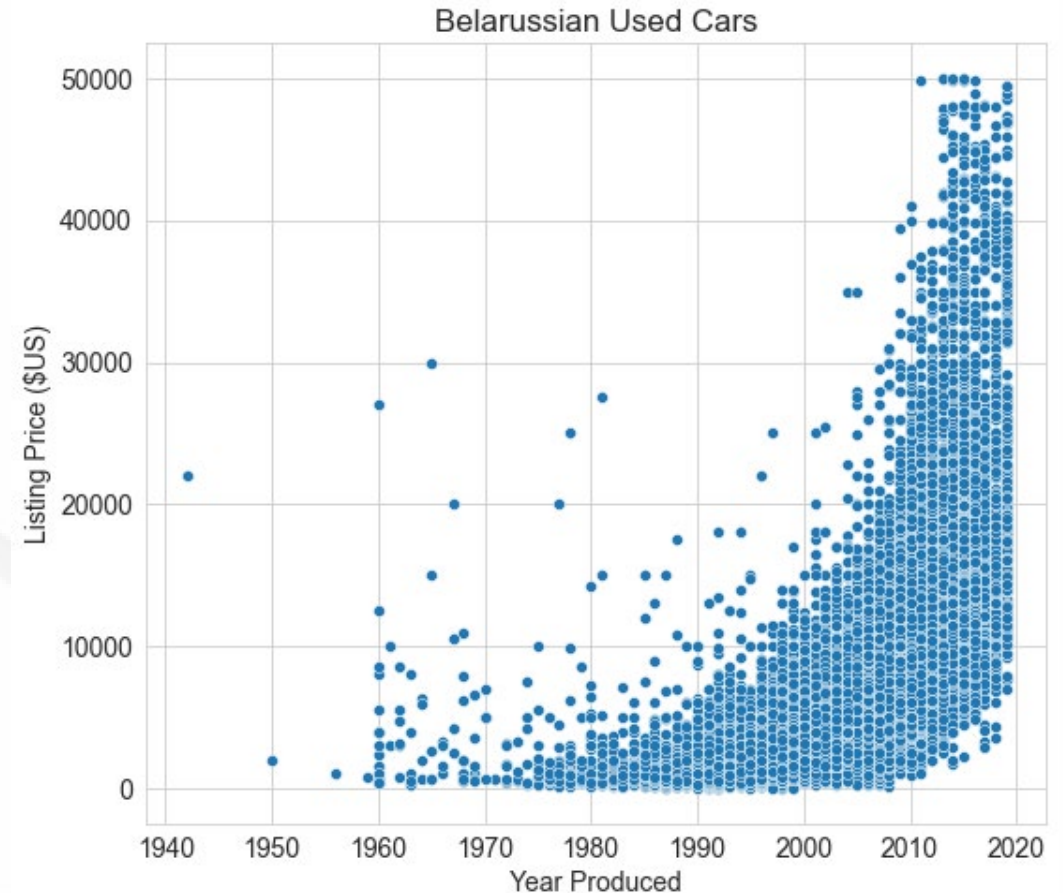
# Box Plot



- Consists of a “Box,” “Whiskers,” and “Outliers”
- Box:
  - Centered on Median
  - Bounded by 1<sup>st</sup>/3<sup>rd</sup> Quartile
  - Contains 50% of Observations
- Whiskers
  - Bounded by farthest observation from the median that is within 1.5x Inter-Quartile Range (IQR) of the median
- Outliers
  - All points beyond 1.5x IQR of the Median

# Scatterplot

- Plots two quantitative data sets against each other
  - If positive-sloped, we say they are positively correlated
  - If negative-sloped, we say they are negatively correlated
- Also useful for finding outliers
- Can be combined with marginal plots or others



# Resources

- Matplotlib Example Gallery
  - <https://matplotlib.org/stable/gallery/index.html>
- Seaborn Example Gallery
  - <https://seaborn.pydata.org/examples/index.html>
- Pandas User Guide
  - [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/index.html](https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html)
- DigitalU
  - <https://digitalu.af.mil/>
- The Visual Display of Quantitative Information by Edward R. Tufte

# Recap

- Intro to visualization in Python
- Rug Plot
- Histogram
- Kernel Density Estimation (KDE) Plot
- Box Plot
- Scatterplot
- Combined plots
- Finalizing your plots