DATA SCIENCE

# Introduction to Data Visualization

Advanced Charts

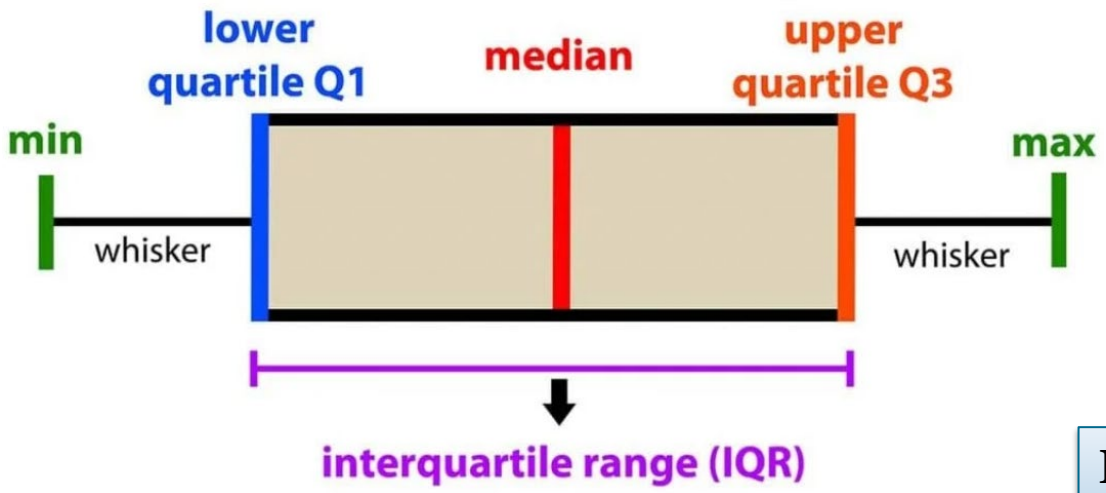


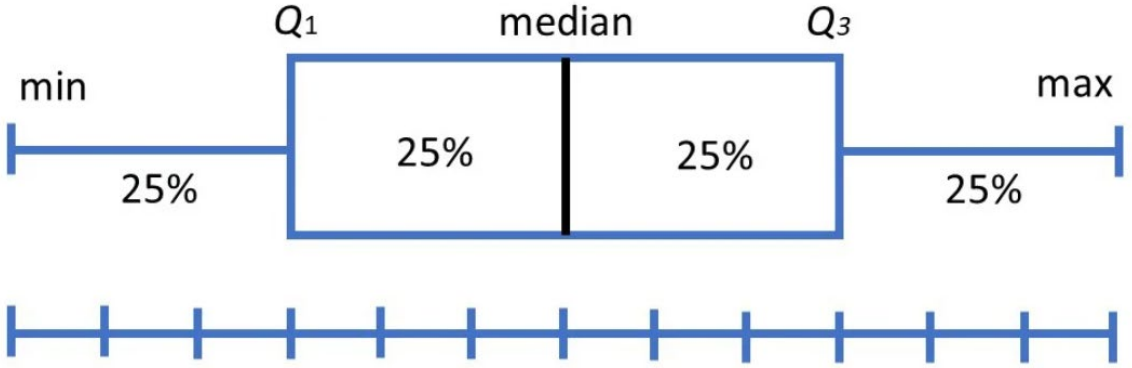
UNC GREENSBORO

## Visualizations of Distribution

1) Box and Whisker Plot

Box plots visually show the distribution of numerical data and skewness by displaying the data quartiles (or percentiles) and averages.



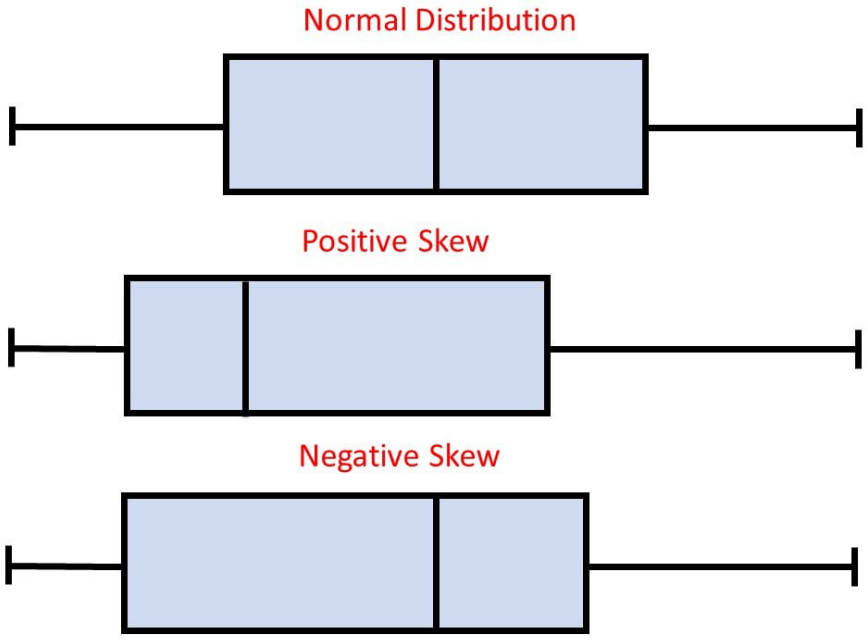




Box plots divide the data into sections containing approximately 25% of the data in that set.

## Visualizations of Distribution

1) Box and Whisker Plot

Box plots visually show the distribution of numerical data and skewness by displaying the data quartiles (or percentiles) and averages.




distribution is symmetric



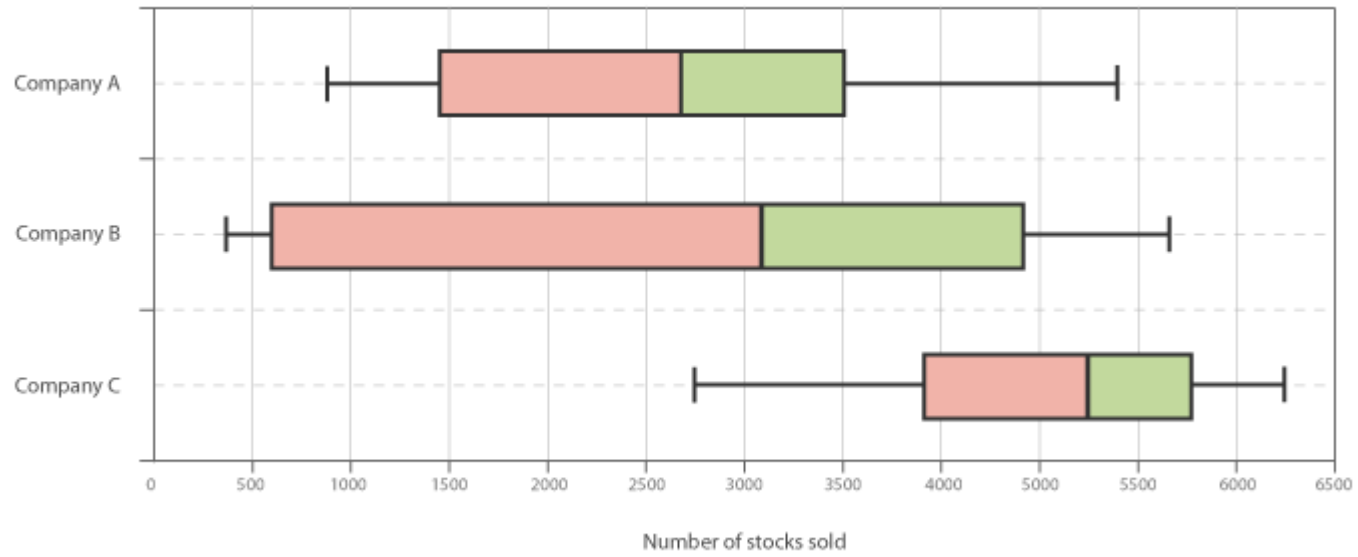

distribution is positively skewed

distribution is negatively skewed

## Visualizations of Distribution

1) Box and Whisker Plot

Box plots visually show the distribution of numerical data and skewness by displaying the data quartiles (or percentiles) and averages.

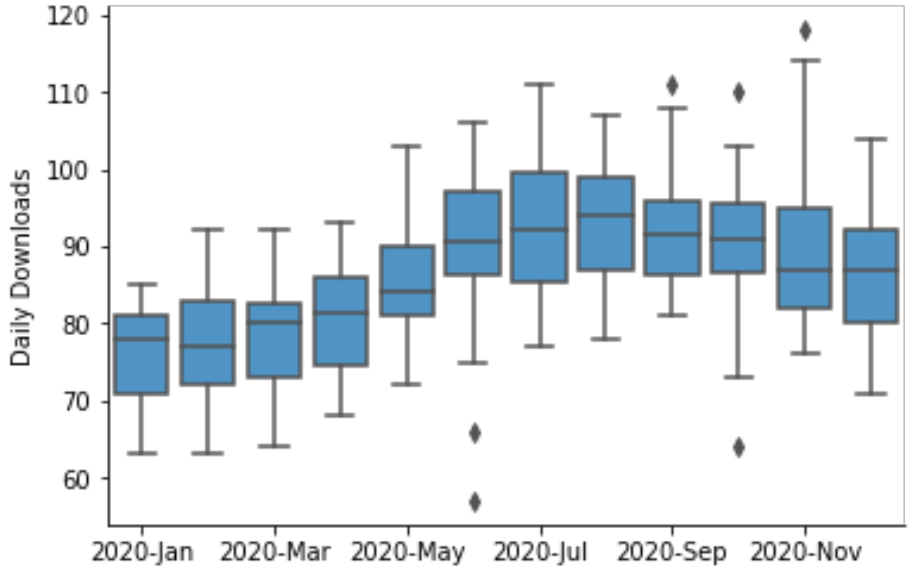


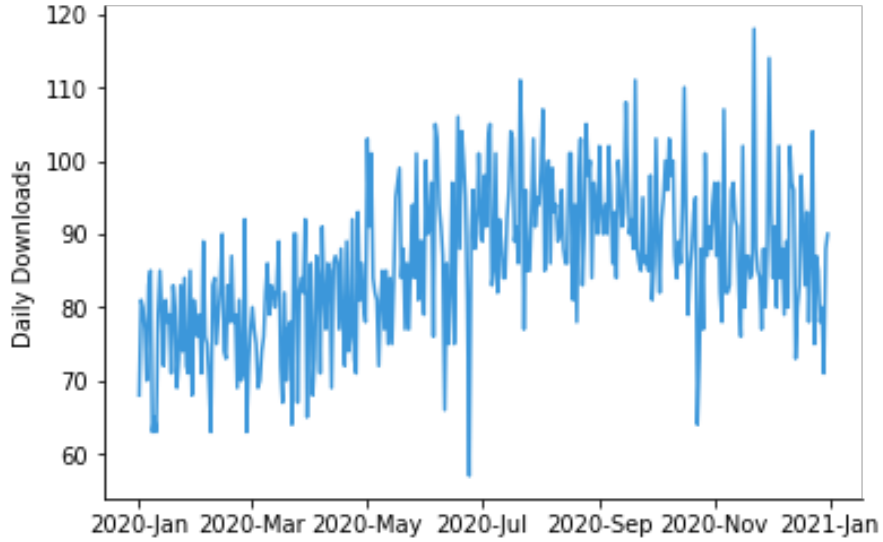
# Visualization

## Visualizations of Distribution

1) Box and Whisker Plot

Box plots visually show the distribution of numerical data and skewness by displaying the data quartiles (or percentiles) and averages.



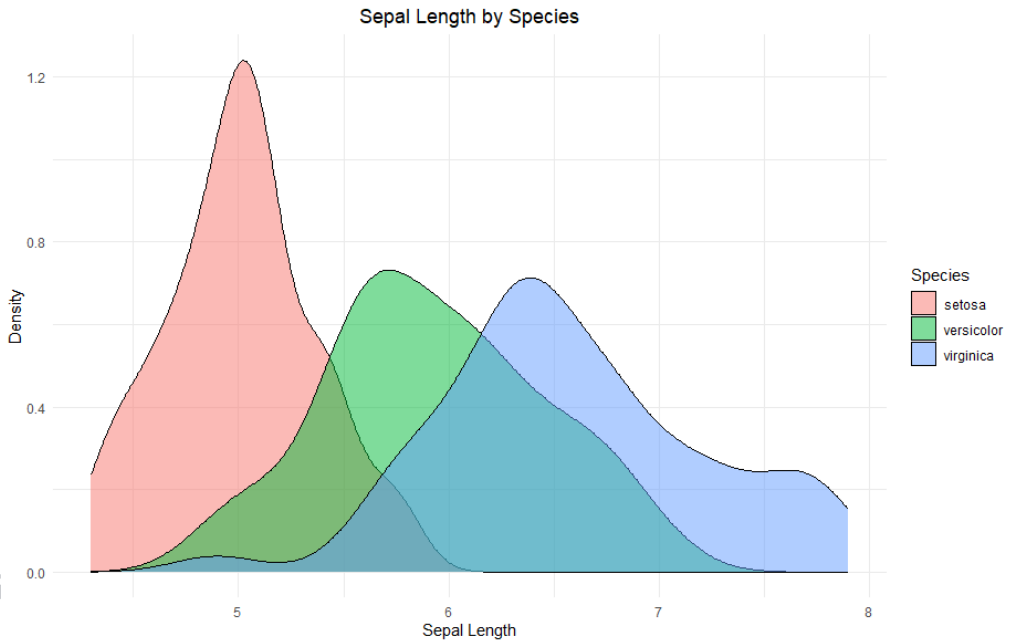



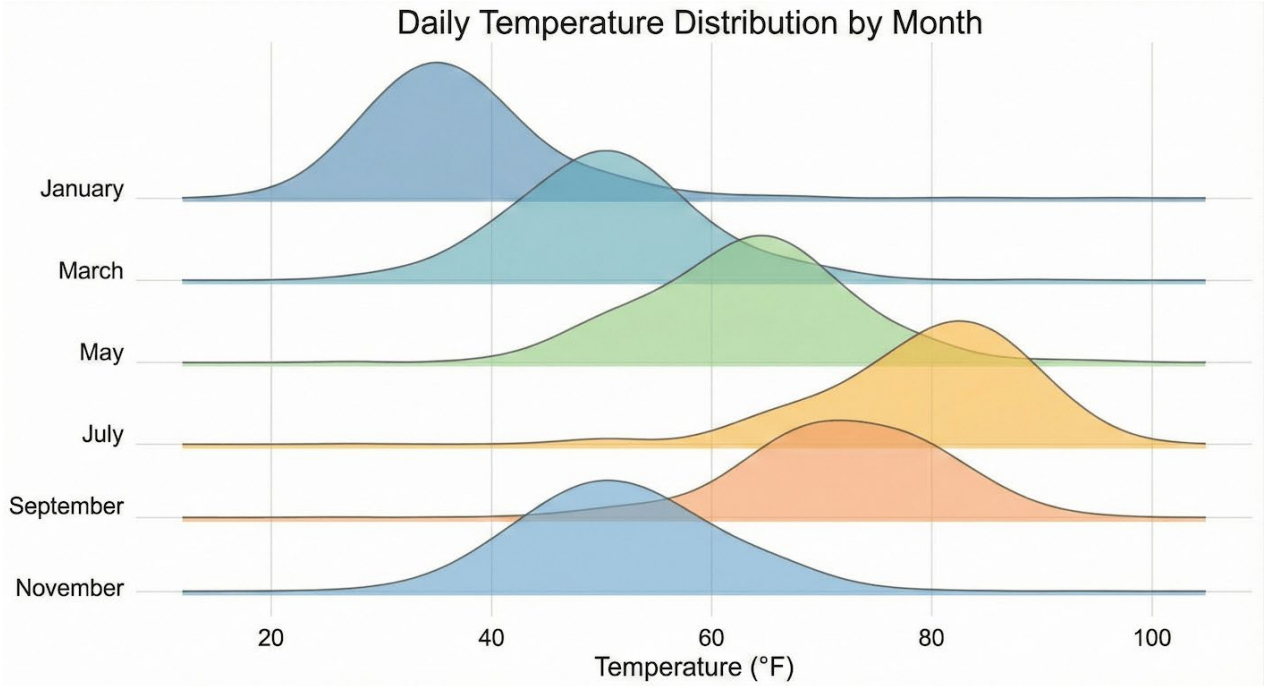

UNC GREENSBORO

# Visualizations of Distribution

## 2) Density Plot

Density plots are used to visualize the **distribution of continuous data**, and are good for showing distribution characteristics, such as shape, central tendency, and spread, making them valuable for understanding data patterns and comparing distributions.

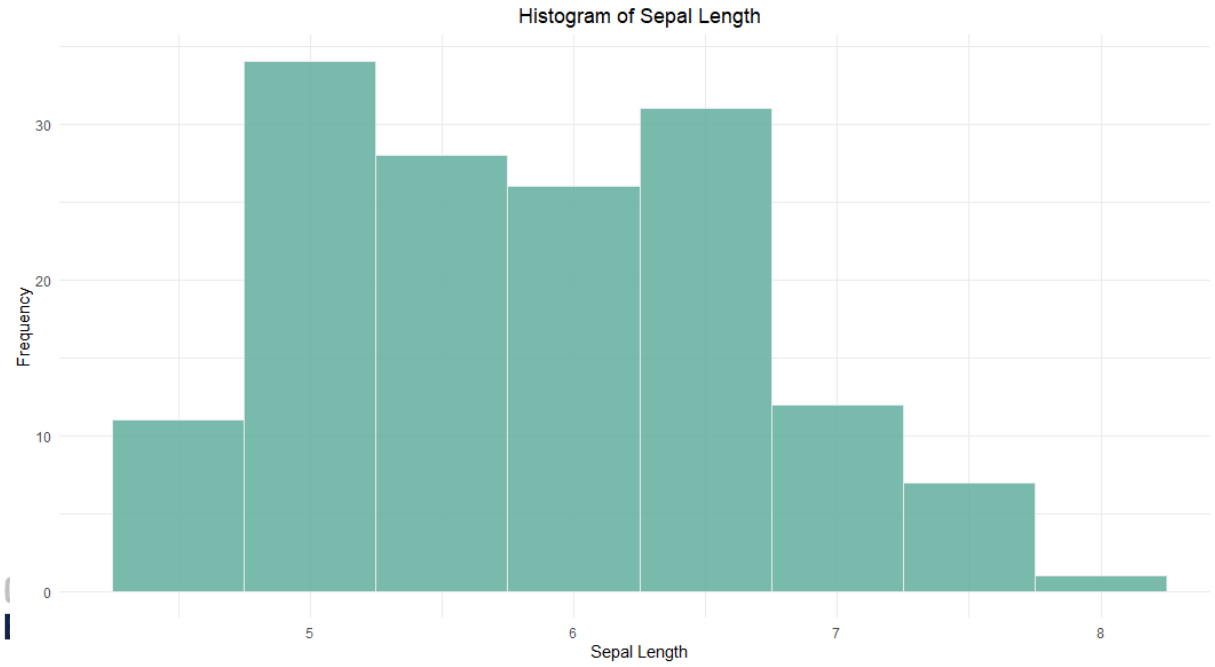


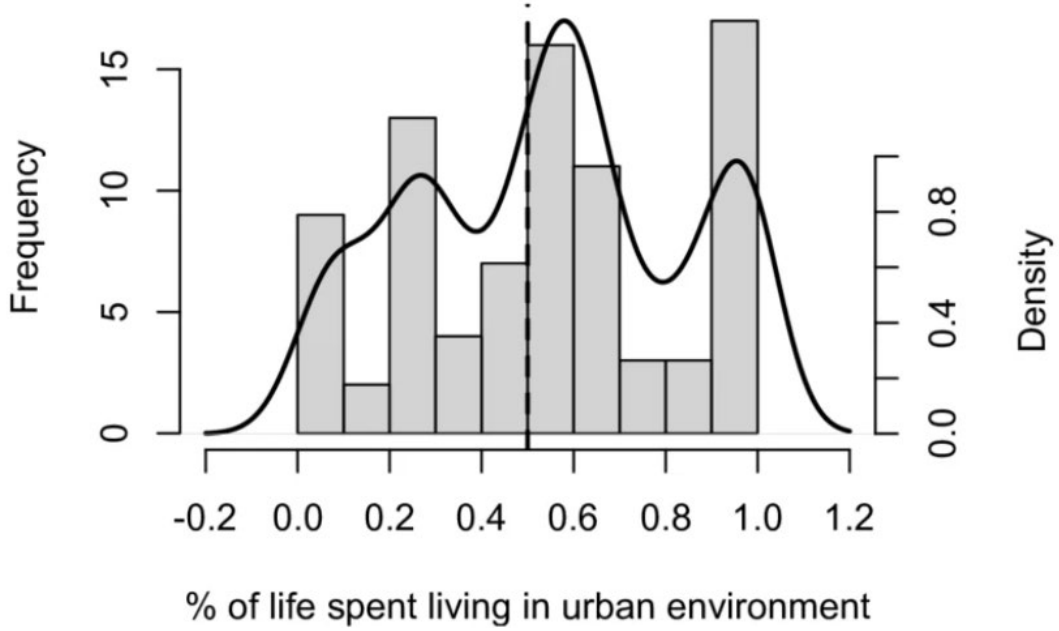



# Visualizations of Distribution

3) Histogram

Histograms are used to visualize the **distribution of continuous data by grouping** it into intervals (bins) and representing the frequency or density of data points within each bin. They are effective for revealing data patterns, central tendencies, and variations.

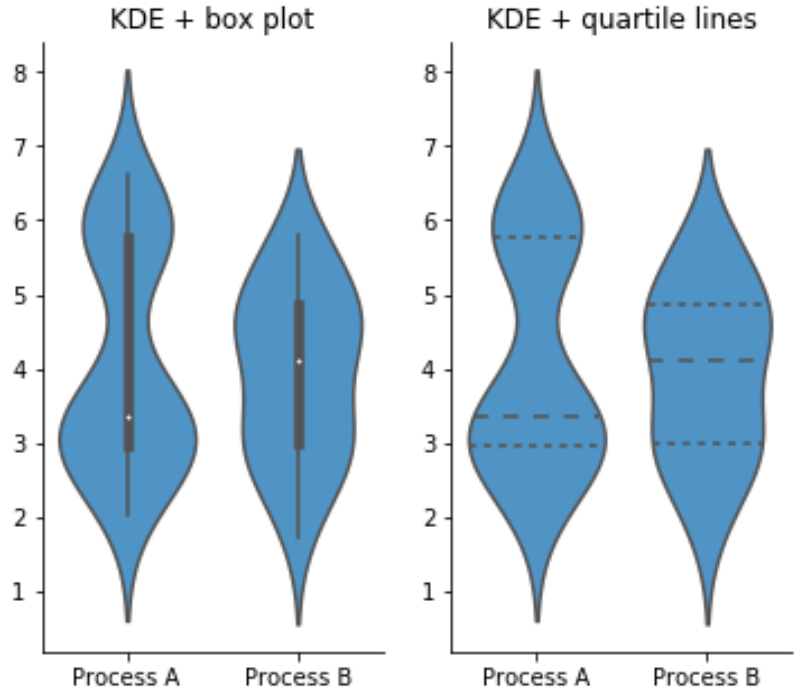


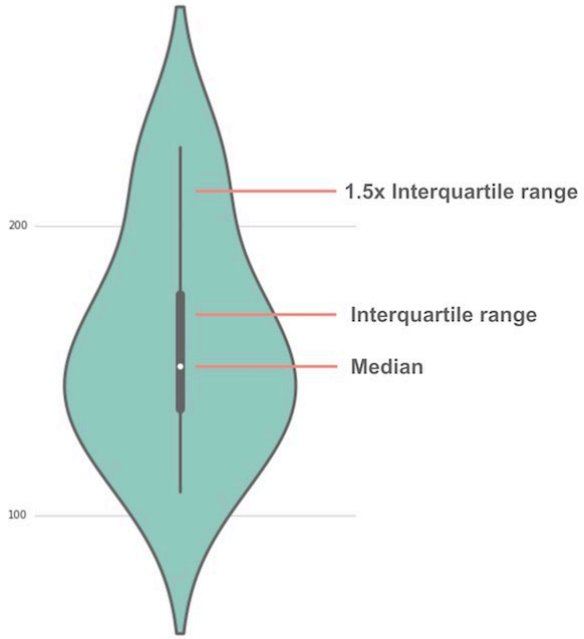



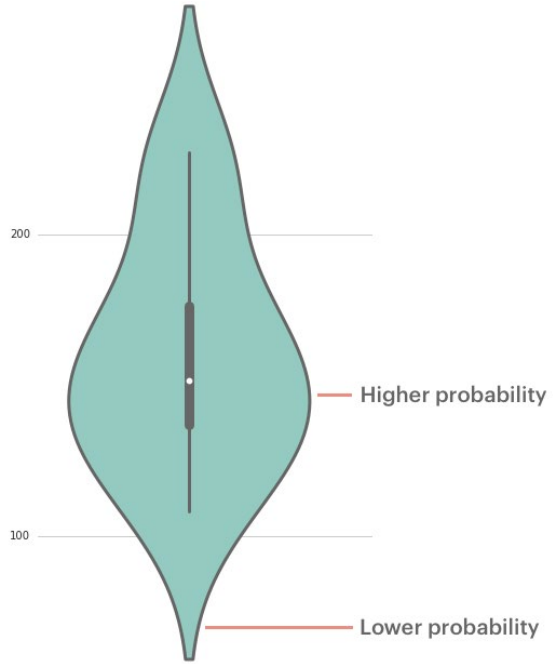
# Visualizations of Distribution

4) violin plot

A violin plot depicts distributions of **numeric data for one or more groups** using density curves. The width of each curve corresponds with the approximate frequency of data points in each region.



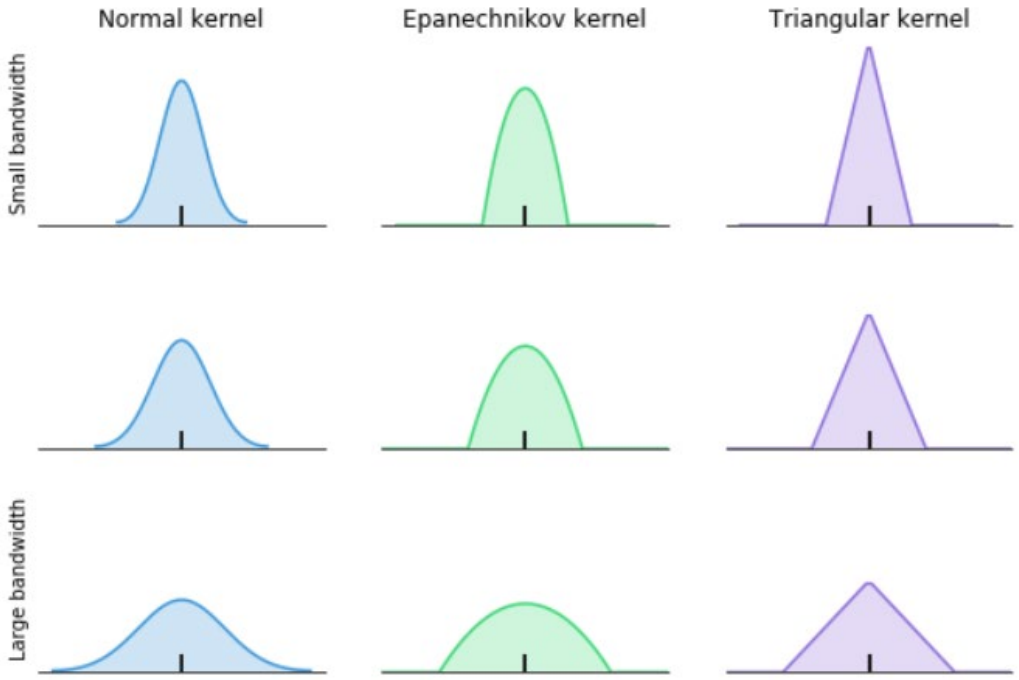



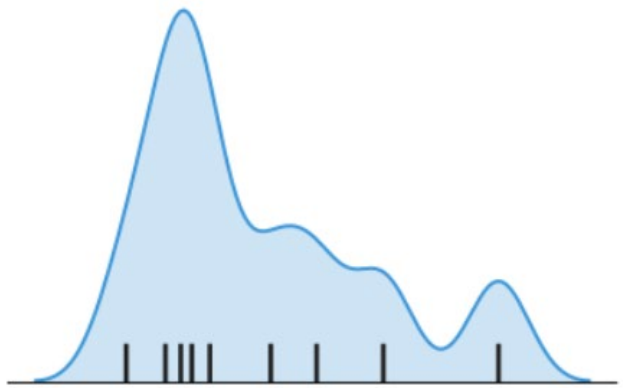



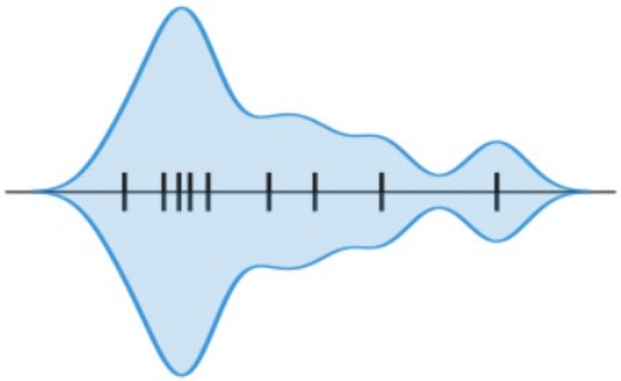
## Visualizations of Distribution

4) violin plot

A violin plot depicts distributions of **numeric data for one or more groups** using density curves. The width of each curve corresponds with the approximate frequency of data points in each region.



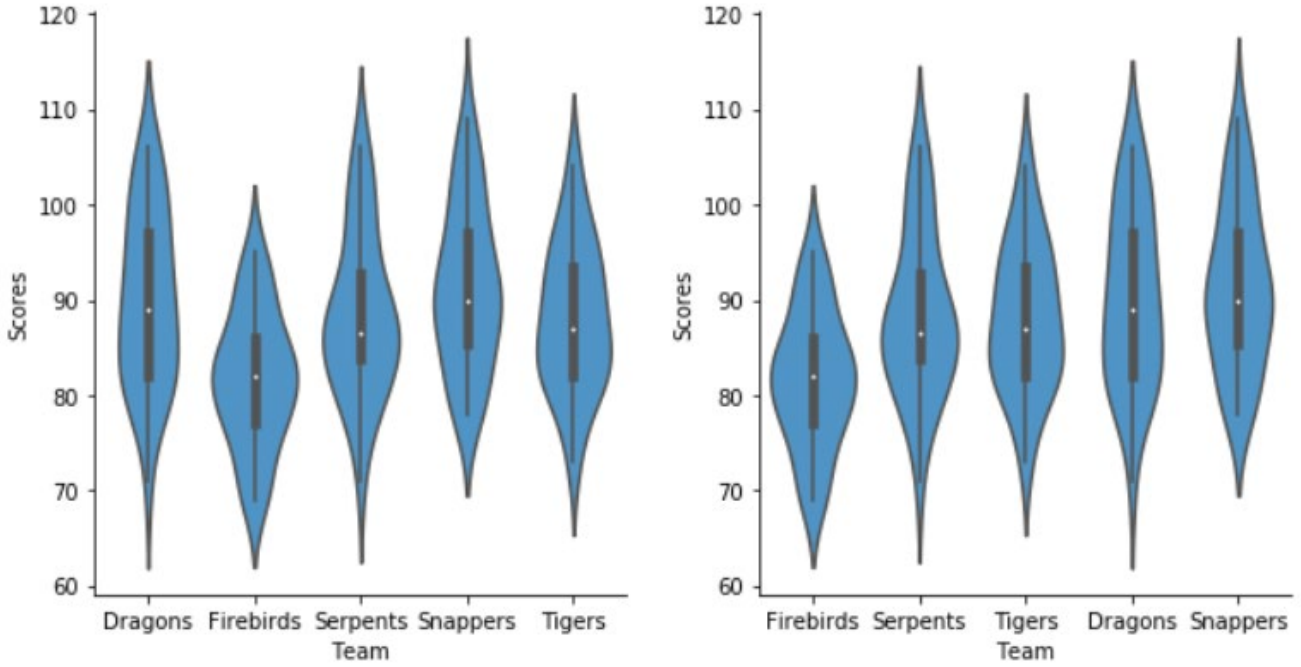


# Visualizations of Distribution

4) violin plot

A violin plot depicts distributions of **numeric data for one or more groups** using density curves. The width of each curve corresponds with the approximate frequency of data points in each region.

Consider the order of groups



# Visualizations of Correlation

5) Correlation Heatmap

A correlation heatmap is a 2D graphical representation of a correlation matrix between multiple variables.

The color intensity of each cell represents the strength of the correlation:

- 1 (or close to 1): Strong positive correlation (dark colors)
- 0: No correlation (neutral colors)
- -1 (or close to -1): Strong negative correlation (light colors)

# Visualizations of Correlation

## 5) Correlation Heatmap

A correlation heatmap is a 2D graphical representation of a correlation matrix between multiple variables.

| | Overall Qual | Total SF | Garage Area | Garage Cars_3.0 | 1st Flr SF | Year Built | Year Remod/Add | Foundation_PConc | Mas Vnr Area | Full Bath_1 | Kitchen Qual_TA | Exter Qual_TA | SalePrice |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Overall Qual** | 1.000000 | 0.534259 | 0.563904 | 0.502657 | 0.477136 | 0.602964 | 0.584654 | 0.571092 | 0.430041 | -0.521553 | -0.568011 | -0.646351 | 0.800207 |
| **Total SF** | 0.534259 | 1.000000 | 0.524145 | 0.399740 | 0.668871 | 0.331811 | 0.300193 | 0.270644 | 0.441001 | -0.418993 | -0.316613 | -0.341000 | 0.716714 |
| **Garage Area** | 0.563904 | 0.524145 | 1.000000 | 0.589214 | 0.498690 | 0.488023 | 0.397731 | 0.393544 | 0.380563 | -0.402050 | -0.365930 | -0.435269 | 0.649897 |
| **Garage Cars_3.0** | 0.502657 | 0.399740 | 0.589214 | 1.000000 | 0.391699 | 0.333050 | 0.303772 | 0.349473 | 0.405799 | -0.295060 | -0.336226 | -0.394001 | 0.619110 |
| **1st Flr SF** | 0.477136 | 0.668871 | 0.498690 | 0.391699 | 1.000000 | 0.323315 | 0.244190 | 0.212511 | 0.386482 | -0.369359 | -0.293941 | -0.318021 | 0.618486 |
| **Year Built** | 0.602964 | 0.331811 | 0.488023 | 0.333050 | 0.323315 | 1.000000 | 0.629116 | 0.666546 | 0.320780 | -0.509293 | -0.478751 | -0.591403 | 0.571849 |
| **Year Remod/Add** | 0.584654 | 0.300193 | 0.397731 | 0.303772 | 0.244190 | 0.629116 | 1.000000 | 0.608503 | 0.204234 | -0.483858 | -0.585228 | -0.590271 | 0.550370 |
| **Foundation_PConc** | 0.571092 | 0.270644 | 0.393544 | 0.349473 | 0.212511 | 0.666546 | 0.608503 | 1.000000 | 0.208299 | -0.500180 | -0.550170 | -0.626157 | 0.529047 |
| **Mas Vnr Area** | 0.430041 | 0.441001 | 0.380563 | 0.405799 | 0.386482 | 0.320780 | 0.204234 | 0.208299 | 1.000000 | -0.229672 | -0.226351 | -0.269285 | 0.503579 |
| **Full Bath_1** | -0.521553 | -0.418993 | -0.402050 | -0.295060 | -0.369359 | -0.509293 | -0.483858 | -0.500180 | -0.229672 | 1.000000 | 0.425653 | 0.496703 | -0.520016 |
| **Kitchen Qual_TA** | -0.568011 | -0.316613 | -0.365930 | -0.336226 | -0.293941 | -0.478751 | -0.585228 | -0.550170 | -0.226351 | 0.425653 | 1.000000 | 0.690116 | -0.540860 |
| **Exter Qual_TA** | -0.646351 | -0.341000 | -0.435269 | -0.394001 | -0.318021 | -0.591403 | -0.590271 | -0.626157 | -0.269285 | 0.496703 | 0.690116 | 1.000000 | -0.600362 |
| **SalePrice** | 0.800207 | 0.716714 | 0.649897 | 0.619110 | 0.618486 | 0.571849 | 0.550370 | 0.529047 | 0.503579 | -0.520016 | -0.540860 | -0.600362 | 1.000000 |

correlation matrix

dataframe.corr()

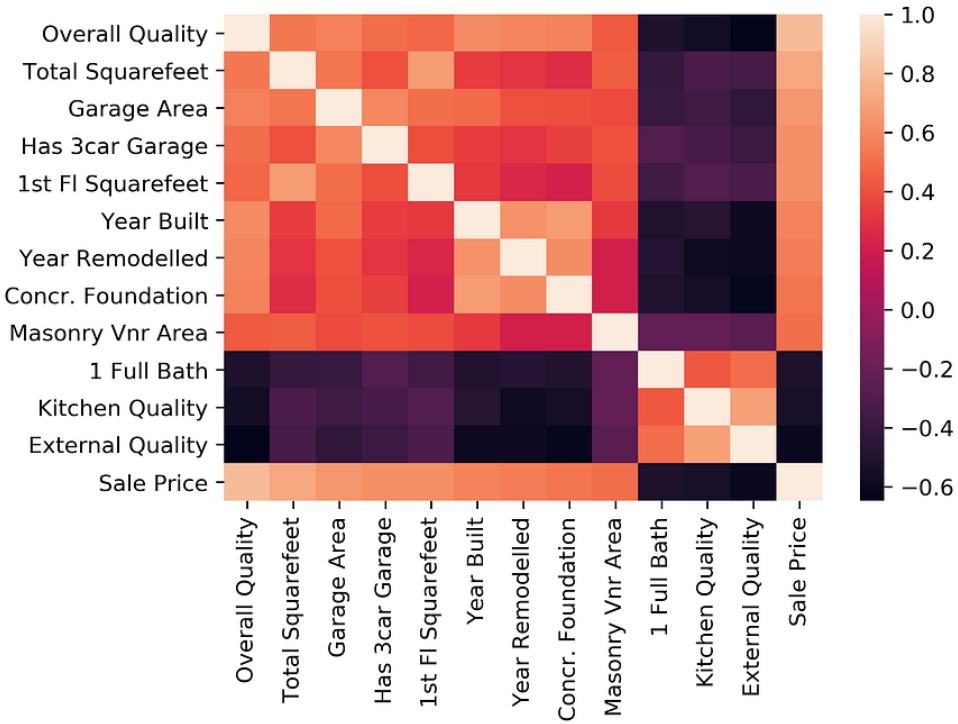
# Visualization

## Visualizations of Correlation

### 5) Correlation Heatmap

A correlation heatmap is a 2D graphical representation of a correlation matrix between multiple variables.

| | Overall Qual | Total SF | Garage Area | Garage Cars_3.0 | 1st Flr SF | Year Built | Year Remod/Add | Foundation_PConc | Mas Vnr Area | Full Bath_1 | Kitchen Qual_TA | Exter Qual_TA | SalePrice |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Overall Qual | 1.000000 | 0.534259 | 0.563904 | 0.502657 | 0.477136 | 0.602964 | 0.584654 | 0.571092 | 0.430041 | -0.521553 | -0.568011 | -0.646351 | 0.800207 |
| Total SF | 0.534259 | 1.000000 | 0.524145 | 0.399740 | 0.668871 | 0.331811 | 0.300193 | 0.270644 | 0.441001 | -0.418993 | -0.316613 | -0.341000 | 0.716714 |
| Garage Area | 0.563904 | 0.524145 | 1.000000 | 0.589214 | 0.498690 | 0.488023 | 0.397731 | 0.393544 | 0.380563 | -0.402050 | -0.365930 | -0.435269 | 0.649897 |
| Garage Cars_3.0 | 0.502657 | 0.399740 | 0.589214 | 1.000000 | 0.391699 | 0.333050 | 0.303772 | 0.349473 | 0.405799 | -0.295060 | -0.336226 | -0.394001 | 0.619110 |
| 1st Flr SF | 0.477136 | 0.668871 | 0.498690 | 0.391699 | 1.000000 | 0.323315 | 0.244190 | 0.212511 | 0.386482 | -0.369359 | -0.293941 | -0.318021 | 0.618486 |
| Year Built | 0.602964 | 0.331811 | 0.488023 | 0.333050 | 0.323315 | 1.000000 | 0.629116 | 0.666546 | 0.320780 | -0.509293 | -0.478751 | -0.591403 | 0.571849 |
| Year Remod/Add | 0.584654 | 0.300193 | 0.397731 | 0.303772 | 0.244190 | 0.629116 | 1.000000 | 0.608503 | 0.204234 | -0.483858 | -0.585228 | -0.590271 | 0.550370 |
| Foundation_PConc | 0.571092 | 0.270644 | 0.393544 | 0.349473 | 0.212511 | 0.666546 | 0.608503 | 1.000000 | 0.208299 | -0.500180 | -0.550170 | -0.626157 | 0.529047 |
| Mas Vnr Area | 0.430041 | 0.441001 | 0.380563 | 0.405799 | 0.386482 | 0.320780 | 0.204234 | 0.208299 | 1.000000 | -0.229672 | -0.226351 | -0.269285 | 0.503579 |
| Full Bath_1 | -0.521553 | -0.418993 | -0.402050 | -0.295060 | -0.369359 | -0.509293 | -0.483858 | -0.500180 | -0.229672 | 1.000000 | 0.425653 | 0.496703 | -0.520016 |
| Kitchen Qual_TA | -0.568011 | -0.316613 | -0.365930 | -0.336226 | -0.293941 | -0.478751 | -0.585228 | -0.550170 | -0.226351 | 0.425653 | 1.000000 | 0.690116 | -0.540860 |
| Exter Qual_TA | -0.646351 | -0.341000 | -0.435269 | -0.394001 | -0.318021 | -0.591403 | -0.590271 | -0.626157 | -0.269285 | 0.496703 | 0.690116 | 1.000000 | -0.600362 |
| SalePrice | 0.800207 | 0.716714 | 0.649897 | 0.619110 | 0.618486 | 0.571849 | 0.550370 | 0.529047 | 0.503579 | -0.520016 | -0.540860 | -0.600362 | 1.000000 |




UNC GREENSBORO

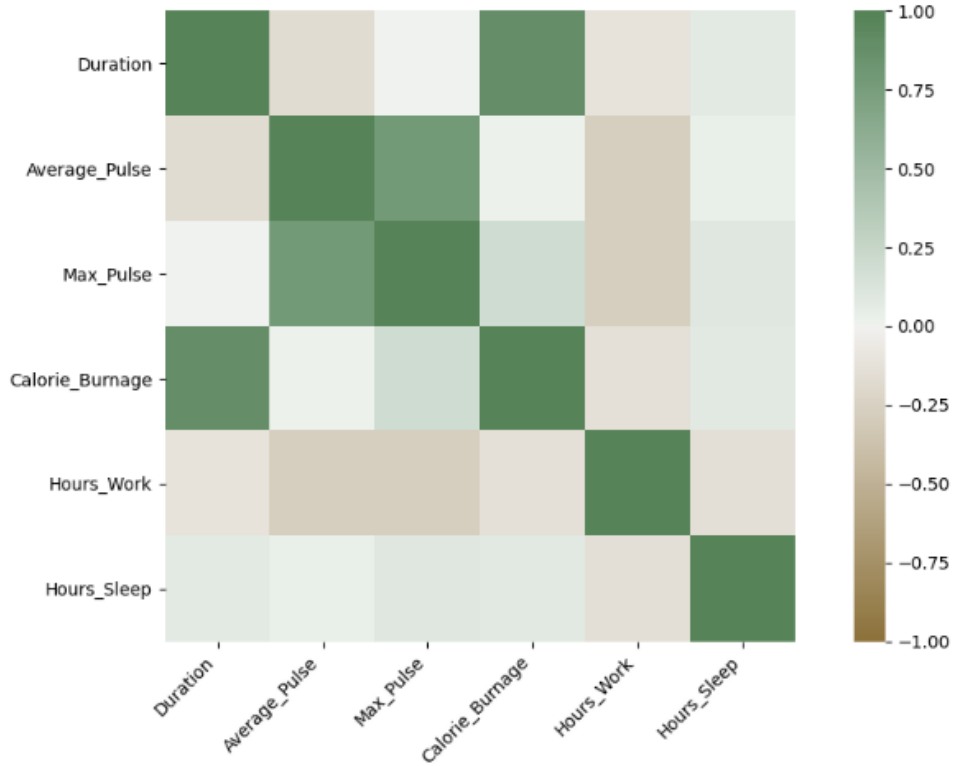
# Visualization

## Visualizations of Correlation

5) Correlation Heatmap

A correlation heatmap is a 2D graphical representation of a correlation matrix between multiple variables.

| | Duration | Average_Pulse | Max_Pulse | Calorie_Burnage |
|---|---|---|---|---|
| Duration | 1.00 | -0.17 | 0.00 | 0.89 |
| Average_Pulse | -0.17 | 1.00 | 0.79 | 0.02 |
| Max_Pulse | 0.00 | 0.79 | 1.00 | 0.20 |
| Calorie_Burnage | 0.89 | 0.02 | 0.20 | 1.00 |




UNC GREENSBORO