Notebook 1: Titanic Dataset 🚢

This notebook is a practical exercise in using **histograms** for data analysis. It focuses on the age distribution of passengers on the Titanic, broken down by gender. Histograms are valuable tools because they show how data is spread out, allowing you to quickly spot patterns and form hypotheses.
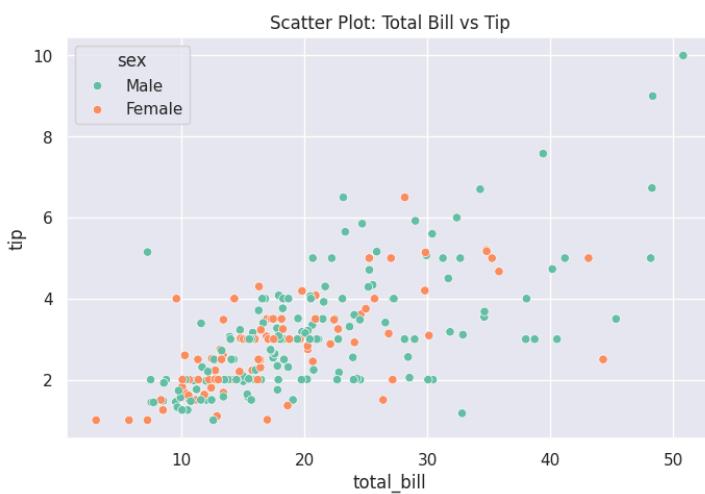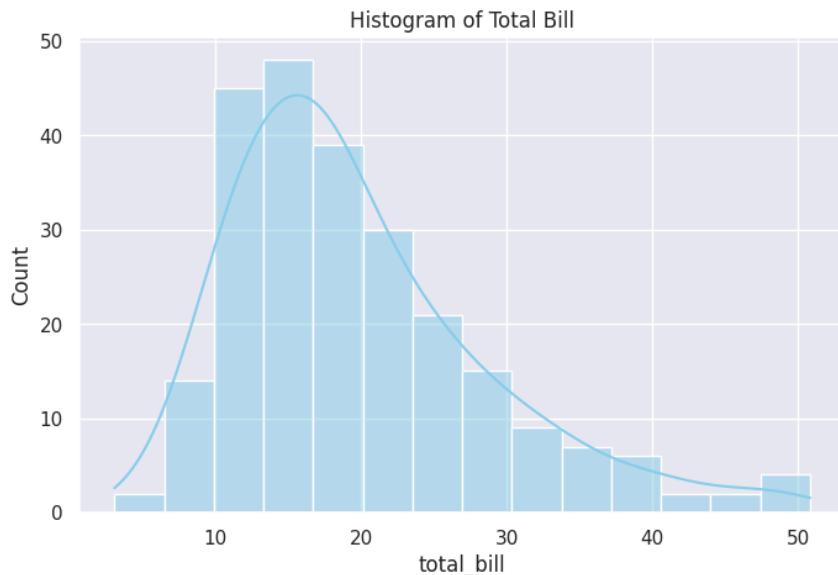
- **Scientific Purpose:** The core purpose is **exploratory data analysis (EDA)**. Before building a complex model to predict survival, a data scientist first needs to understand the dataset's basic demographics. This visualization helps answer fundamental questions like: "What was the age profile of passengers?" and "Were there more men or women on board?" The insights gained here, such as the age skew and gender imbalance, are crucial for guiding further, more advanced analysis.
- **Key Learnings from the Histograms:**
  - **Gender Imbalance:** The histograms immediately show that there were significantly more male passengers than female passengers. This is a critical finding that would influence any further analysis on survival rates.
  - **Age Distribution:** Both graphs show a **right-skewed distribution**, meaning most passengers were young adults (20s and 30s), with a smaller number of older passengers. A right-skewed distribution has a long "tail" on the right side, pulled by a few high values.
  - **Subtle Differences:** The female histogram shows a slightly higher proportion of children and elderly women. This observation, combined with historical knowledge of the "women and children first" policy, suggests a potential link between age, gender, and survival outcomes.
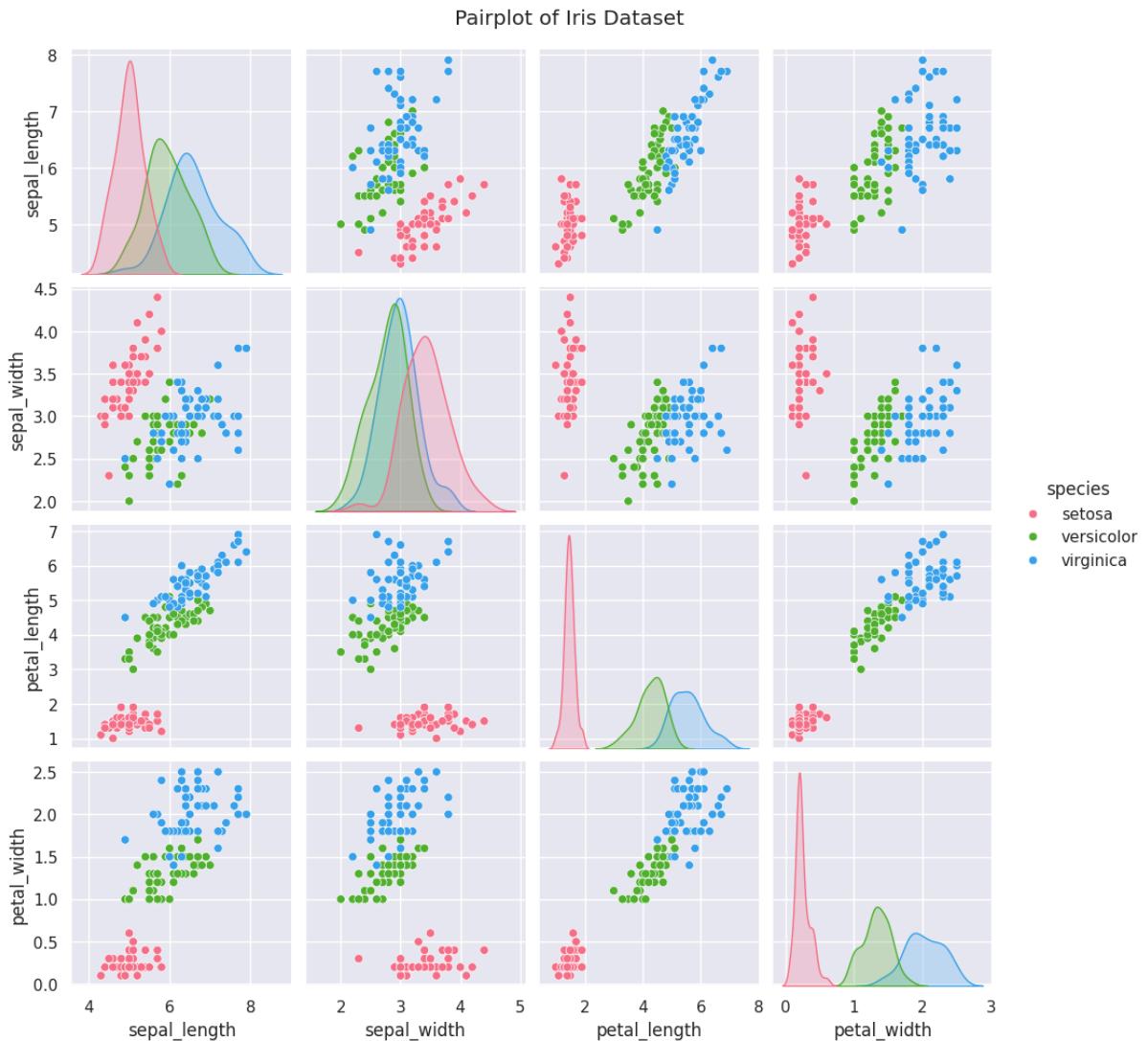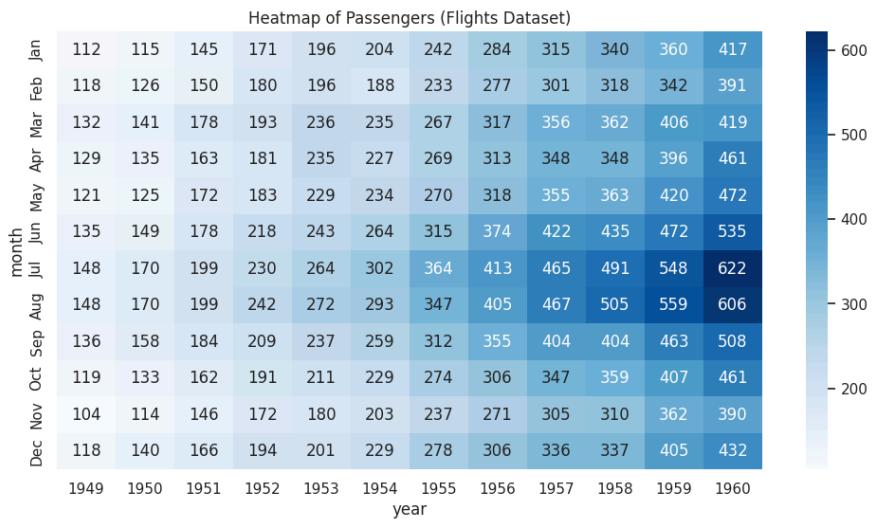
Notebook 2: Data Visualization Lab
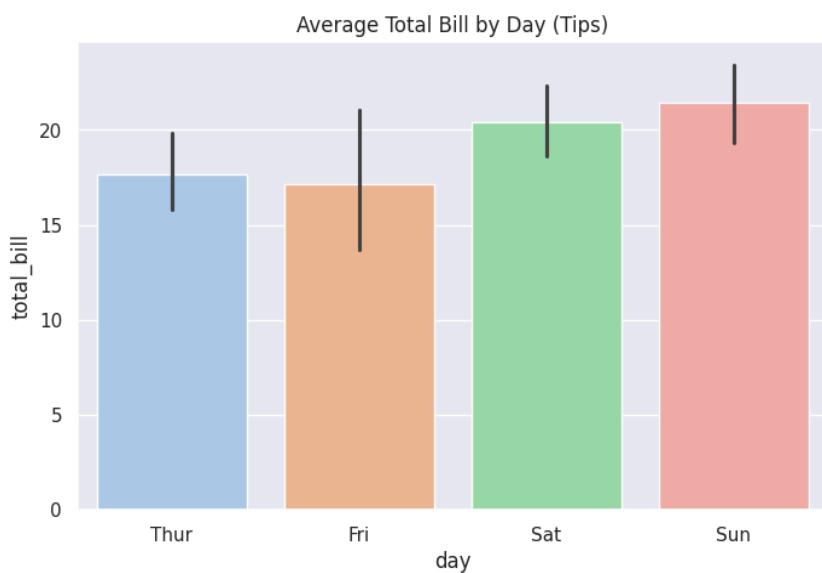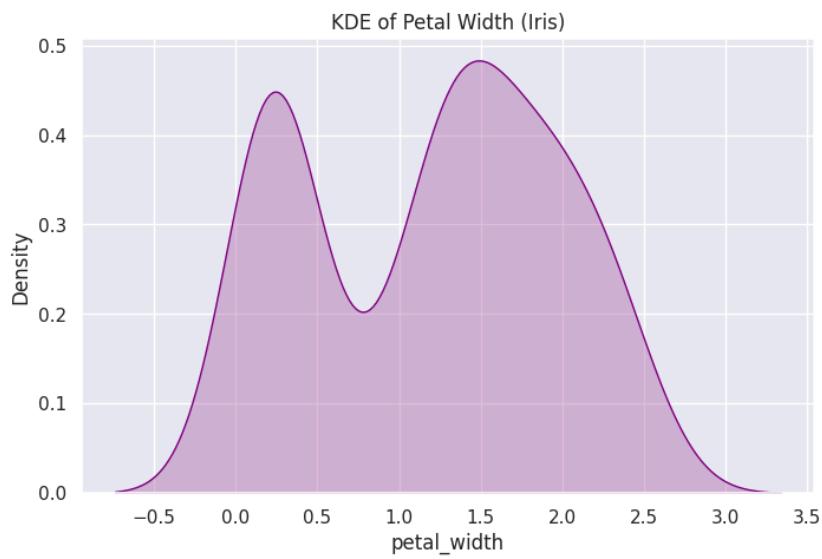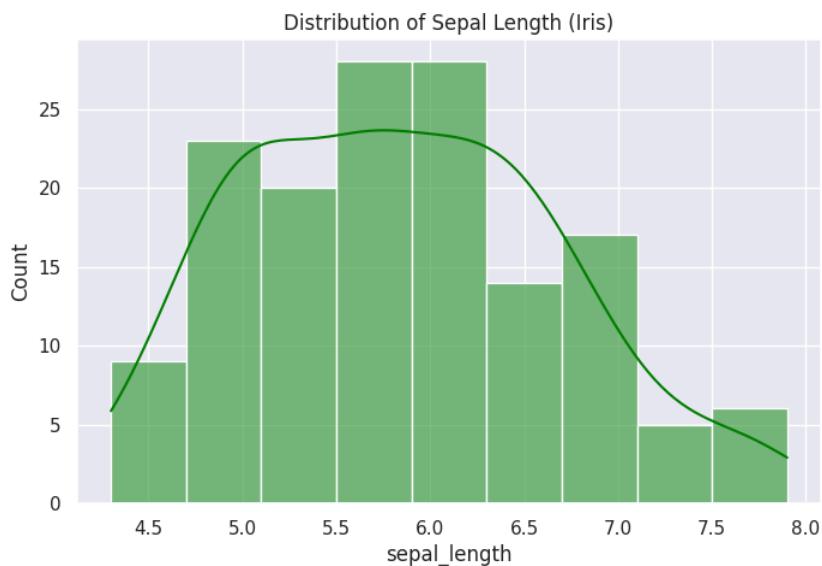
This notebook is a broader lab assignment that showcases various Seaborn plot types to analyze different datasets. The scientific purpose is to teach the principle of **choosing the right visualization for the right question**, as each plot serves a unique purpose in data exploration.
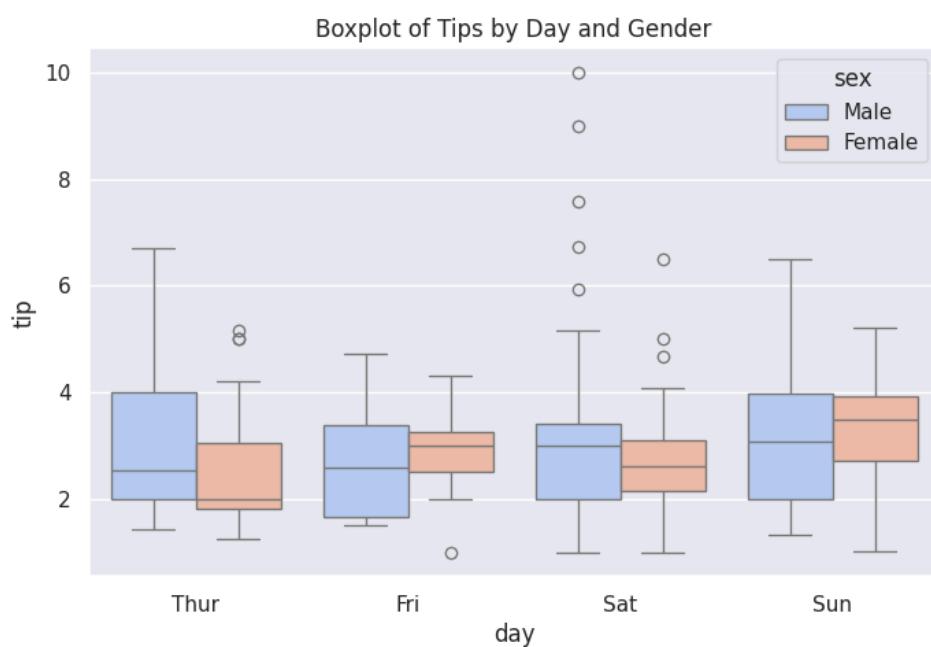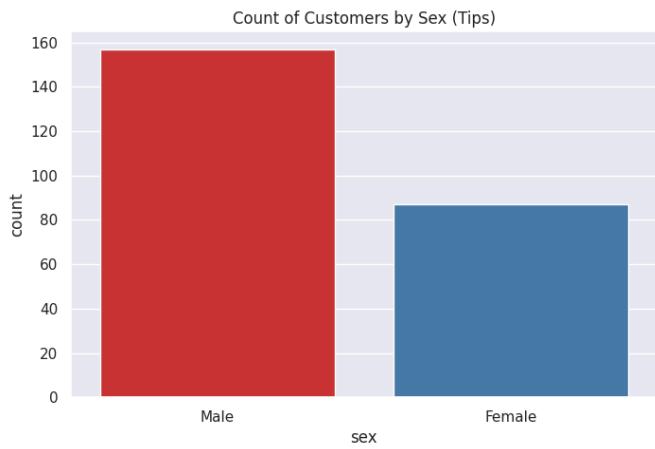
- **Scientific Goals by Plot Type:**
  - **Histograms & Scatterplots:** The `tips` dataset demonstrates both **univariate** (single variable) and **bivariate** (two variables) analysis. The histogram of `total_bill` shows its right-skewed distribution. The scatterplot of `total_bill` vs. `tip` reveals a **positive linear relationship**, meaning larger bills are associated with larger tips.
  - **Heatmaps:** The `flights` dataset uses a heatmap to visualize time-series data. This plot is excellent for showing patterns over time. The heatmap reveals a clear **upward trend** in air travel from 1949 to 1960 and a distinct **seasonal pattern**, with summer months consistently having the most passengers.
  - **Pairplots:** The `iris` dataset uses a **pairplot**, a powerful tool for **multivariate analysis** (exploring relationships among many variables at once). This plot is

particularly useful for classification problems. It immediately shows that the **petal length and petal width** are the most effective features for distinguishing between the three species of iris flowers, as they form clear, separate clusters. This insight is essential for building a machine learning model to classify flowers.


Histogram of Total Bill

Scatter Plot: Total Bill vs Tip

Heatmap of Passengers (Flights Dataset)

month	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
Jan	112	115	145	171	196	204	242	284	315	340	360	417
Feb	118	126	150	180	196	188	233	277	301	318	342	391
Mar	132	141	178	193	236	235	267	317	356	362	406	419
Apr	129	135	163	181	235	227	269	313	348	348	396	461
May	121	125	172	183	229	234	270	318	355	363	420	472
Jun	135	149	178	218	243	264	315	374	422	435	472	535
Jul	148	170	199	230	264	302	364	413	465	491	548	622
Aug	148	170	199	242	272	293	347	405	467	505	559	606
Sep	136	158	184	209	237	259	312	355	404	404	463	508
Oct	119	133	162	191	211	229	274	306	347	359	407	461
Nov	104	114	146	172	180	203	237	271	305	310	362	390
Dec	118	140	166	194	201	229	278	306	336	337	405	432

Pairplot of Iris Dataset

species
- setosa
- versicolor
- virginica

Distribution of Sepal Length (Iris)

KDE of Petal Width (Iris)

Average Total Bill by Day (Tips)

Count of Customers by Sex (Tips)


Boxplot of Tips by Day and Gender

Violin Plot of Petal Length by Species (Iris)
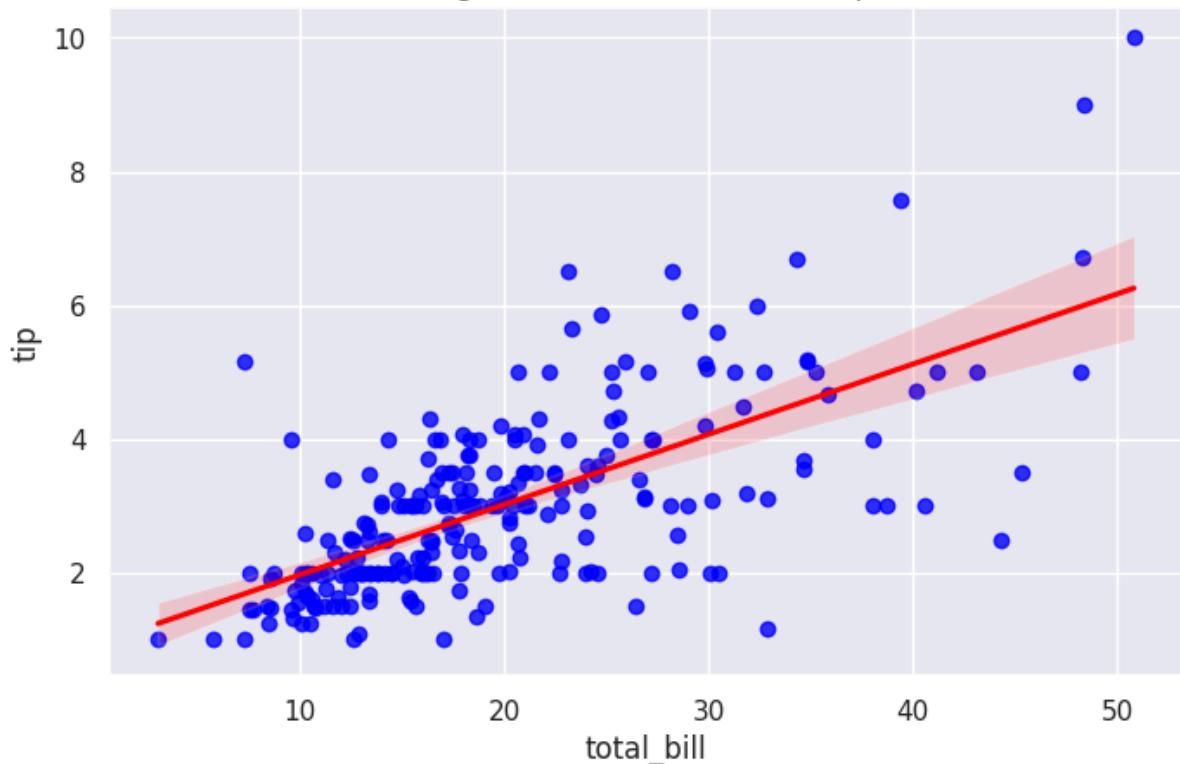
Swarmplot of Petal Width by Species (Iris)
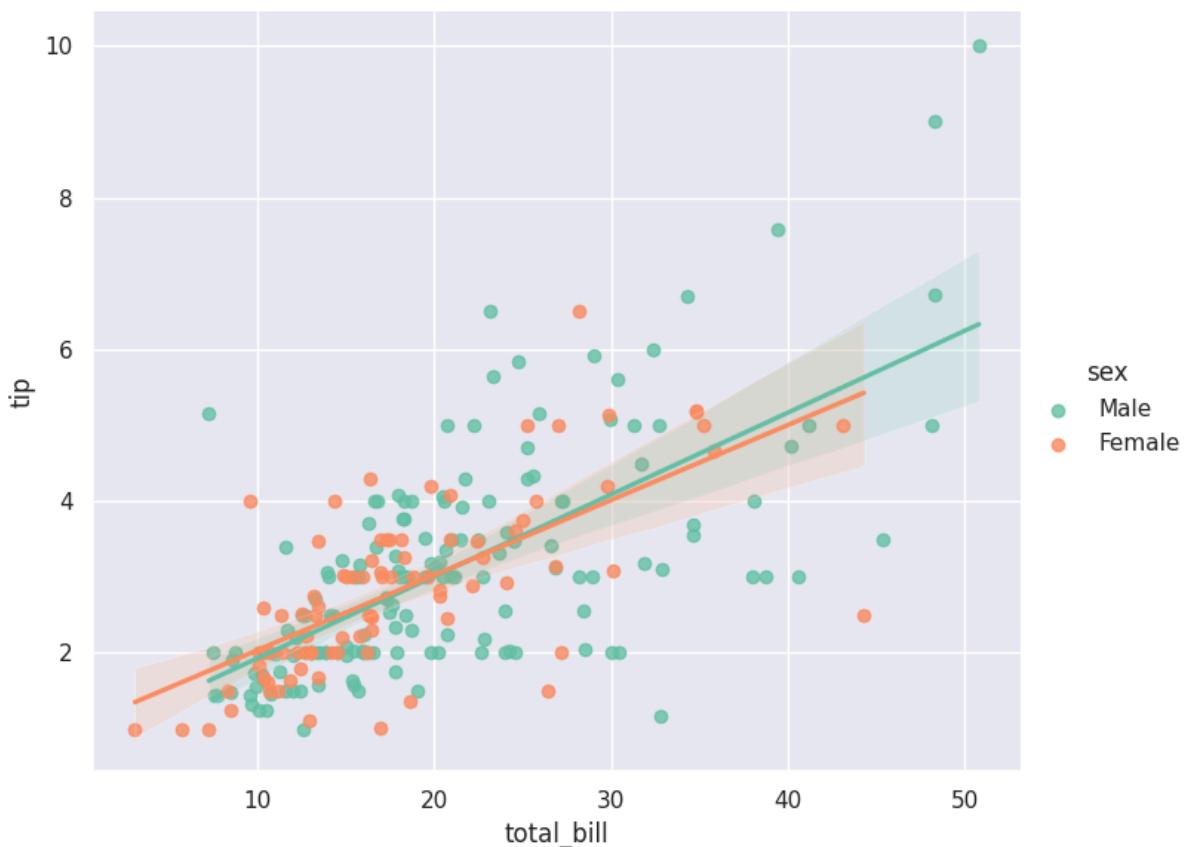
Clustermap of Flights Passengers
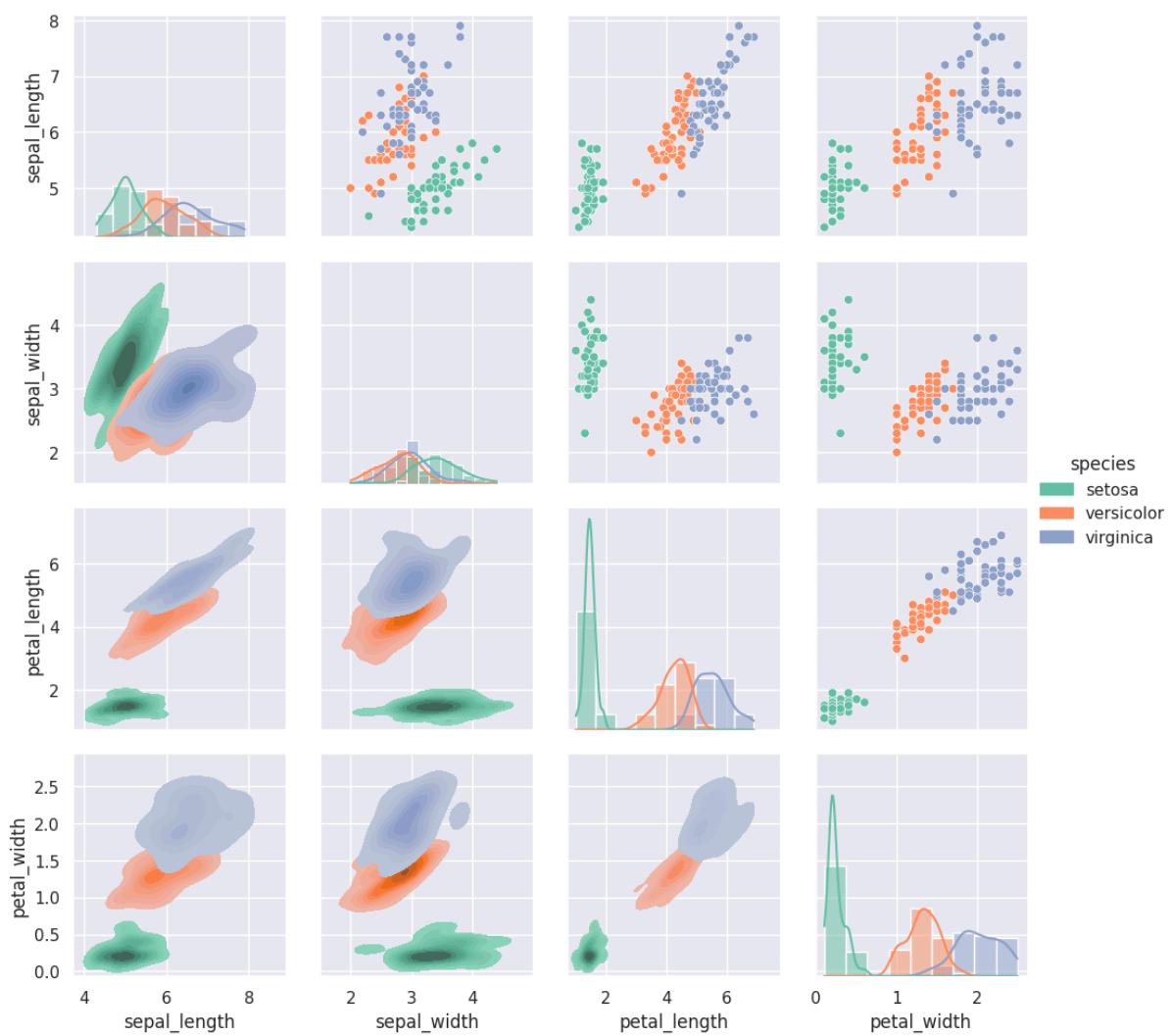
Regression Plot: Total Bill vs Tip
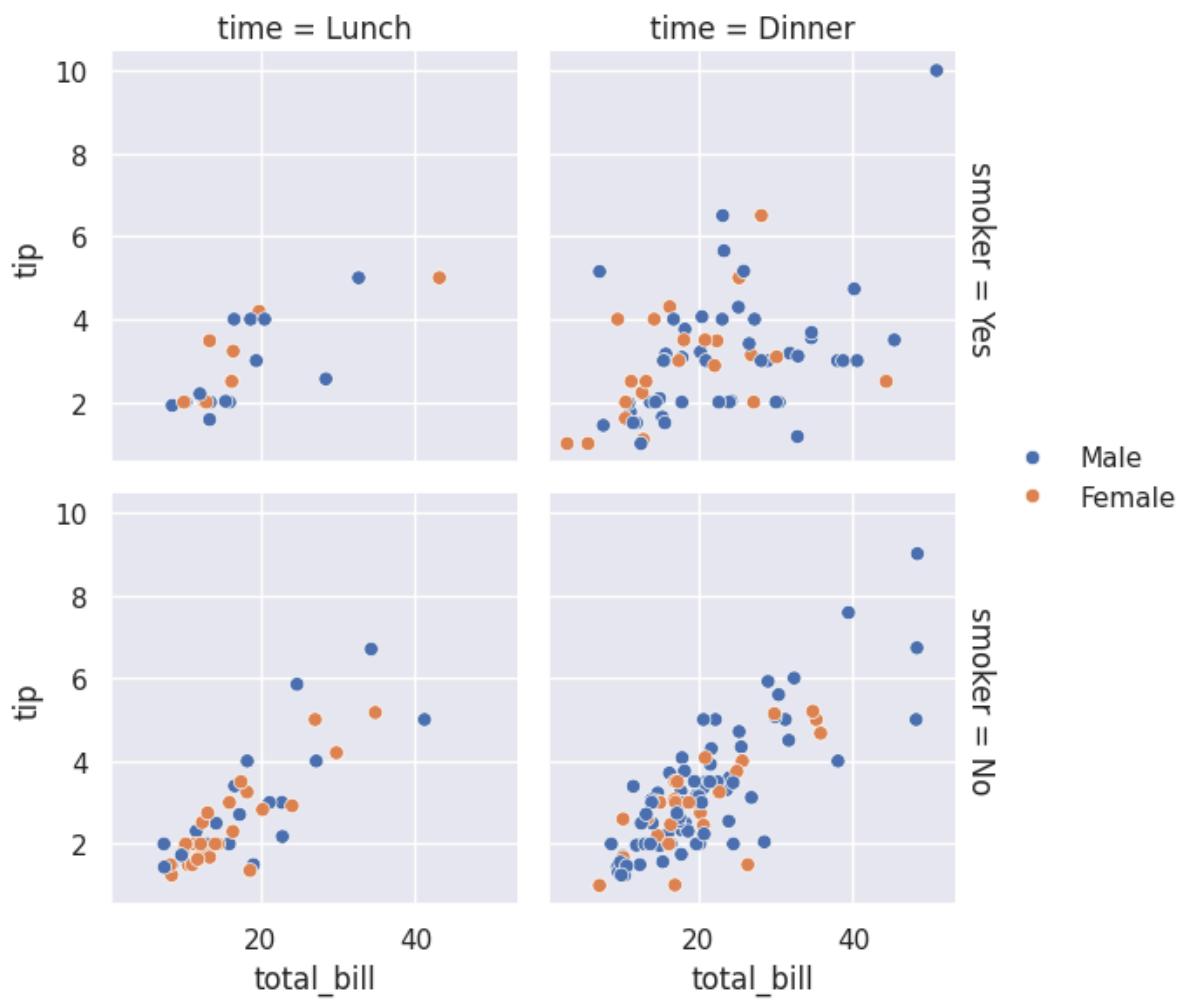
LMPlot: Total Bill vs Tip by Sex
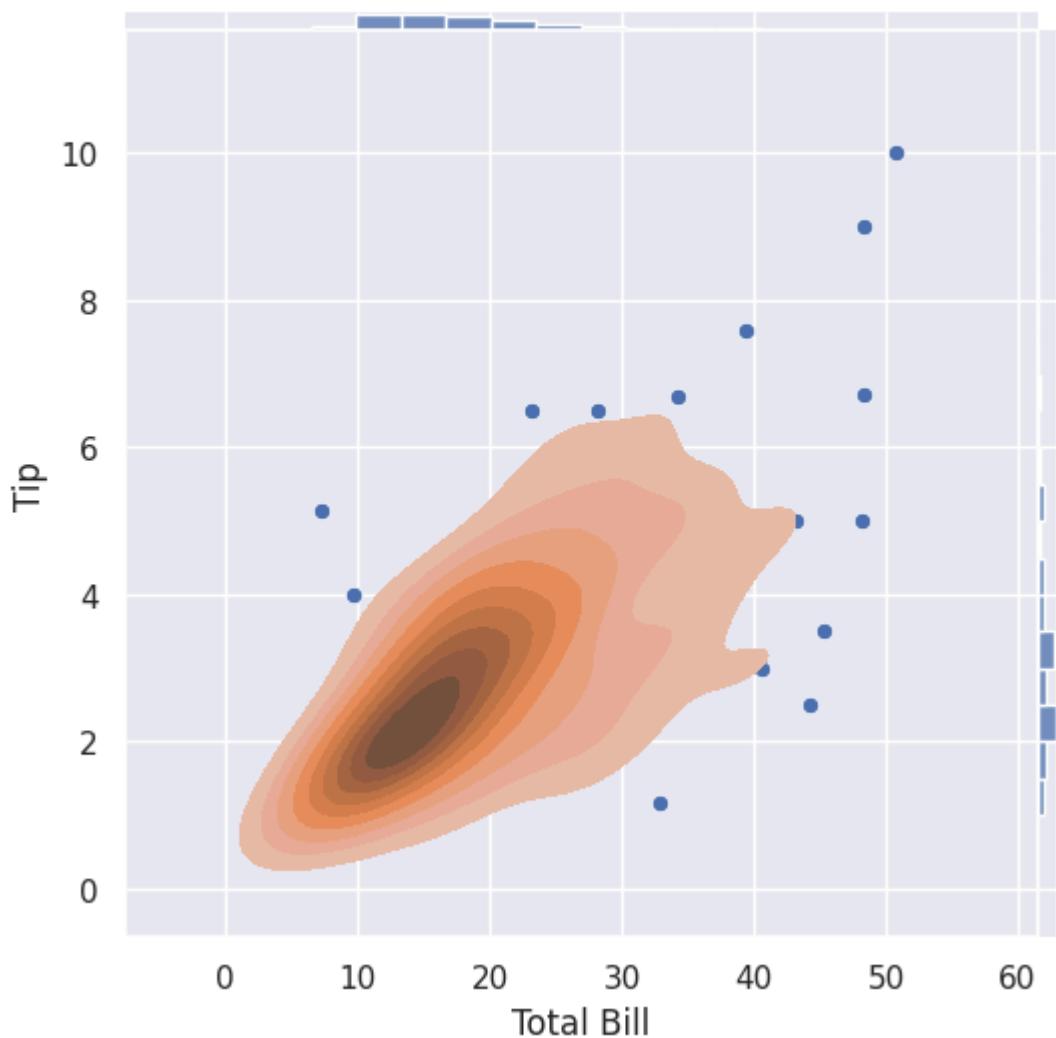
PairGrid of Iris Dataset
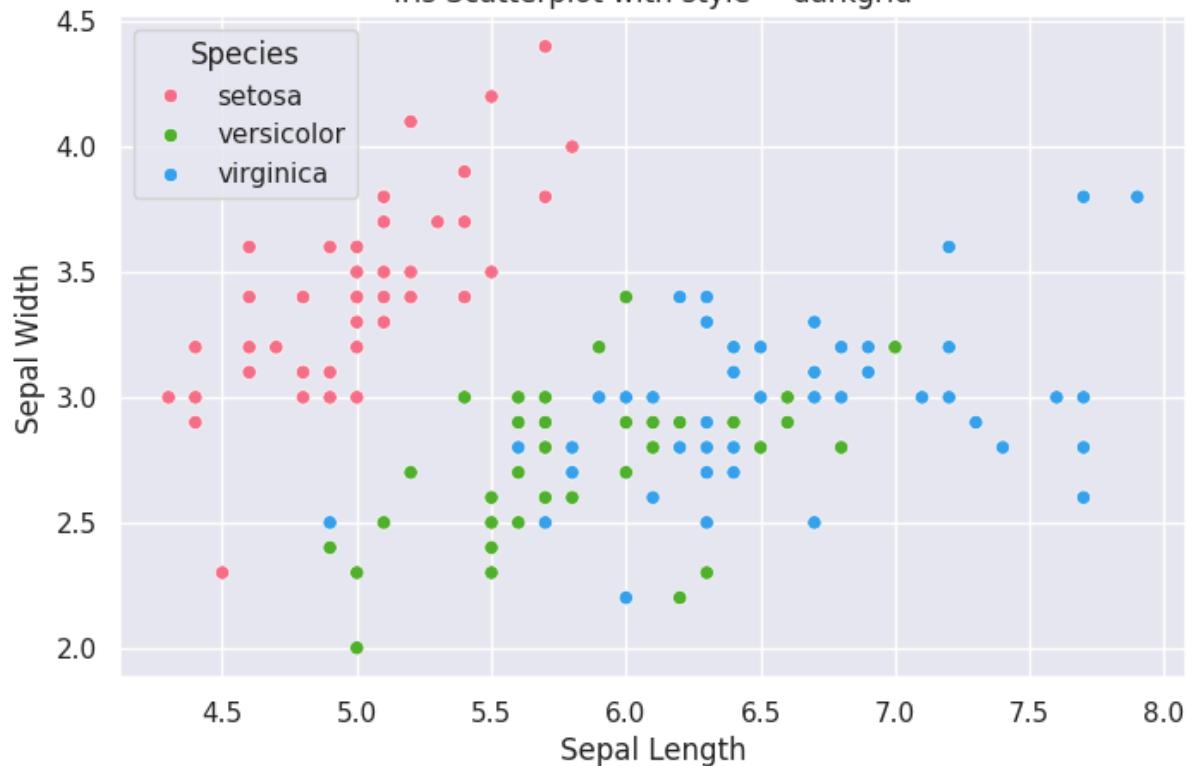
FacetGrid of Tips Dataset by Time and Smoker
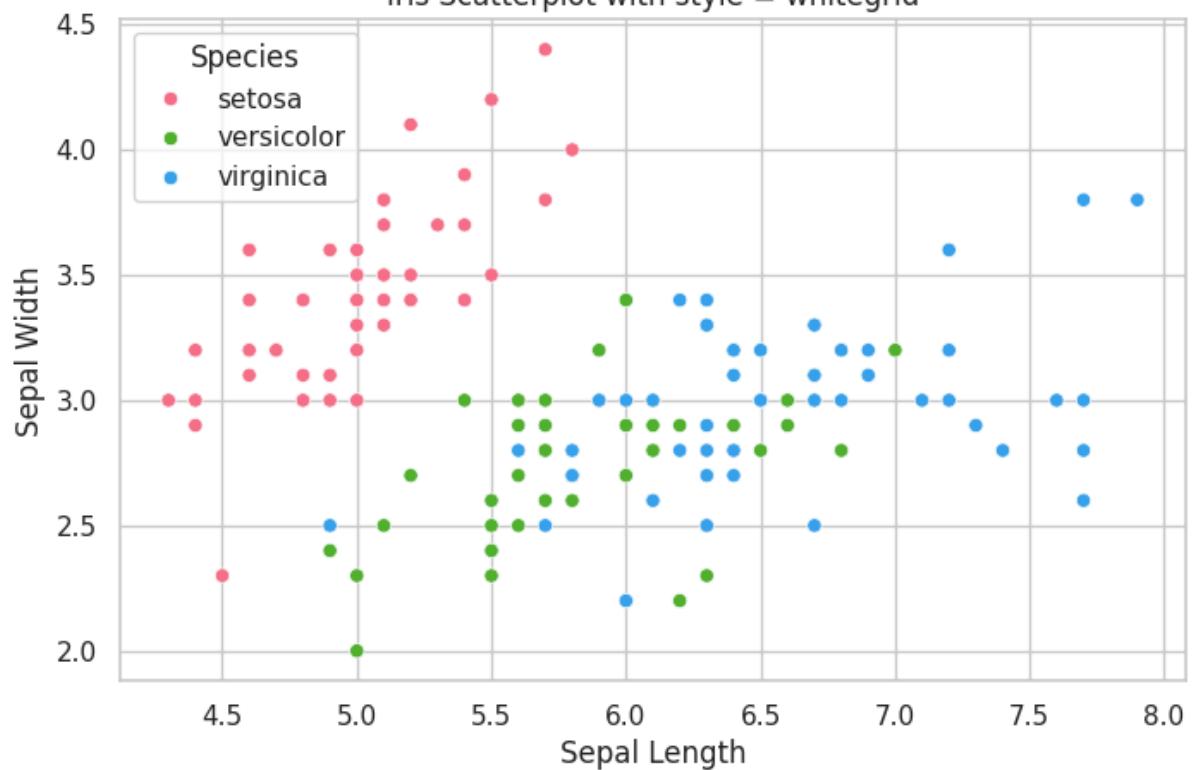
JointGrid: Total Bill vs Tip (Tips Dataset)
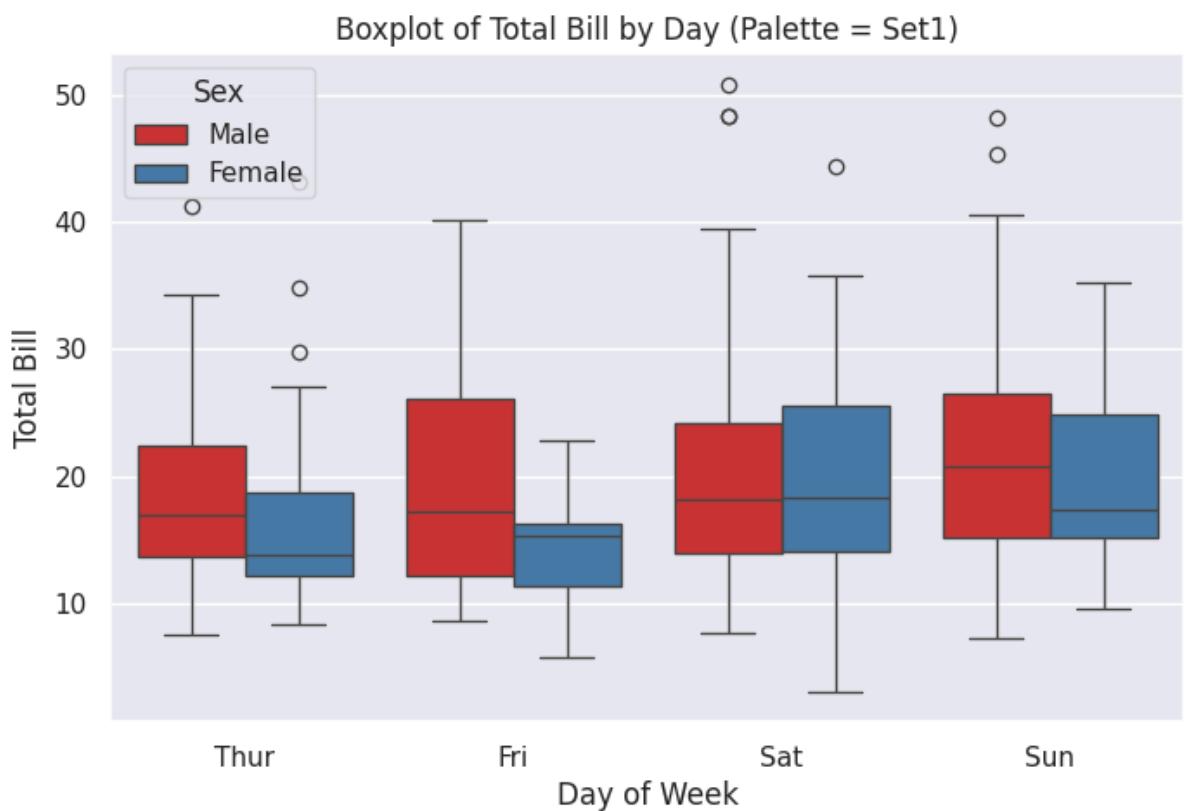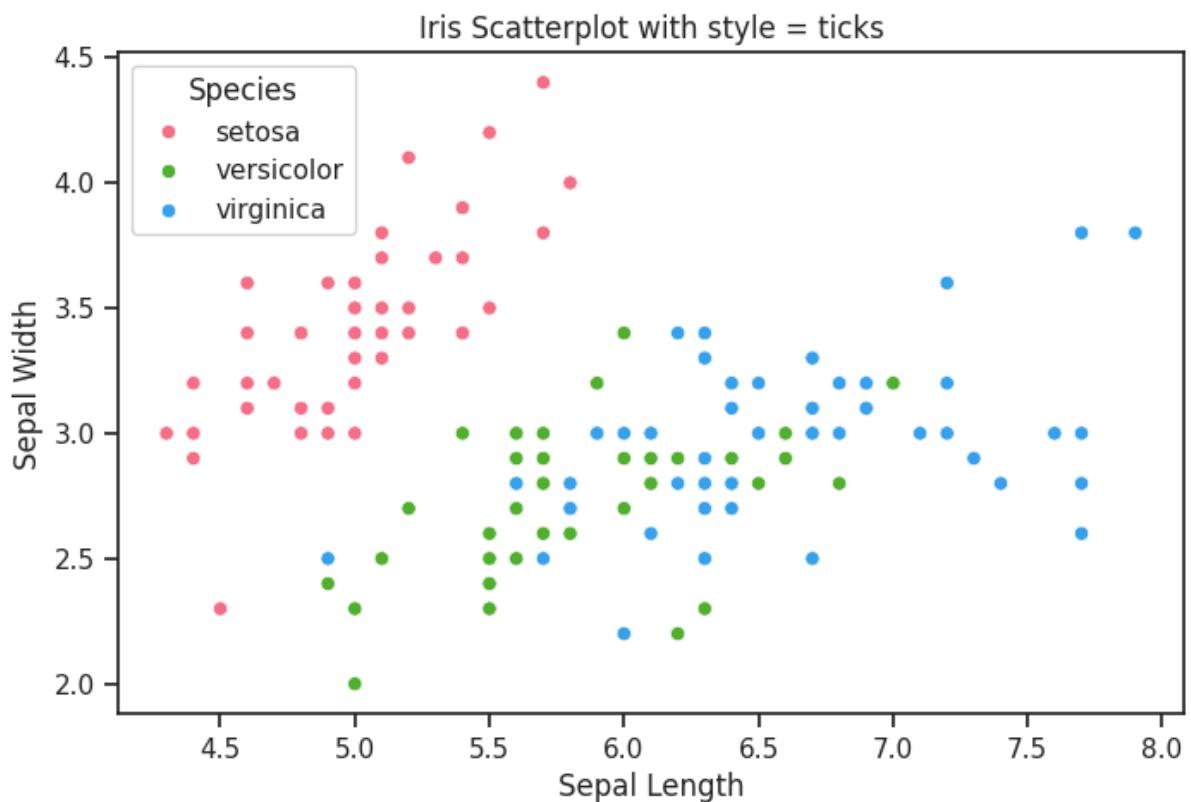
Iris Scatterplot with style = darkgrid
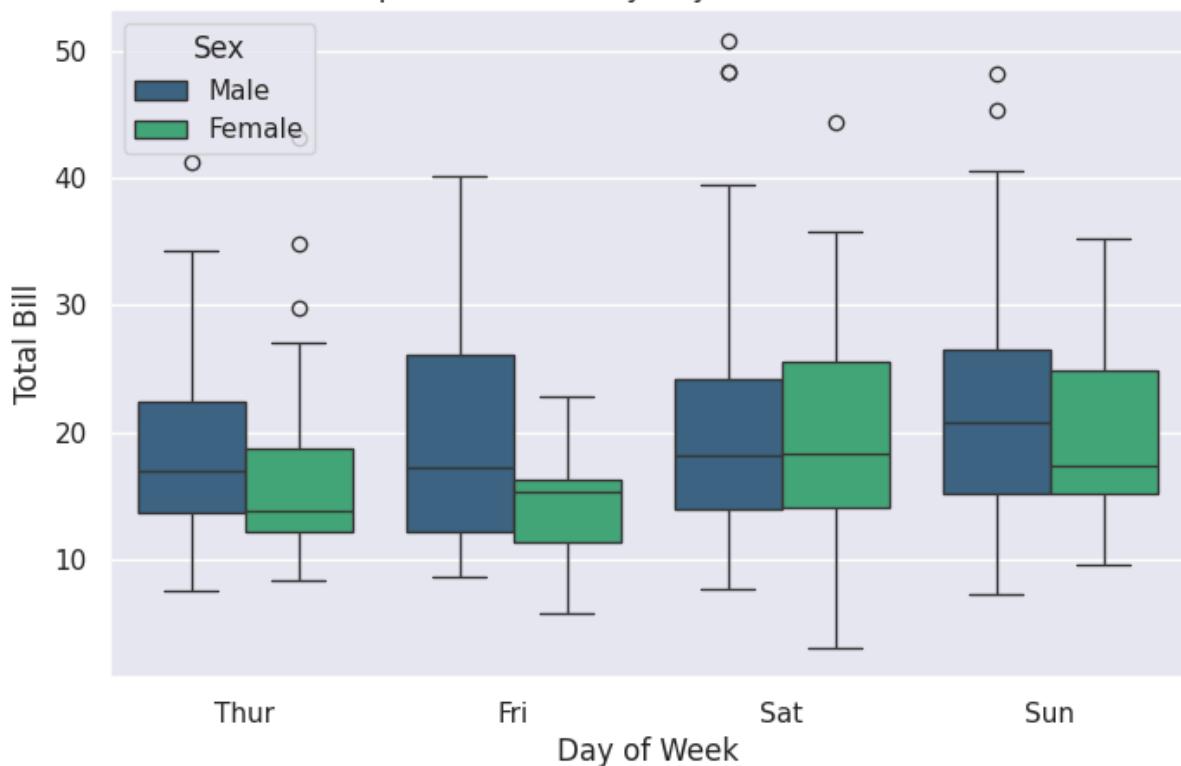
Iris Scatterplot with style = whitegrid

Iris Scatterplot with style = ticks

Boxplot of Total Bill by Day (Palette = Set1)
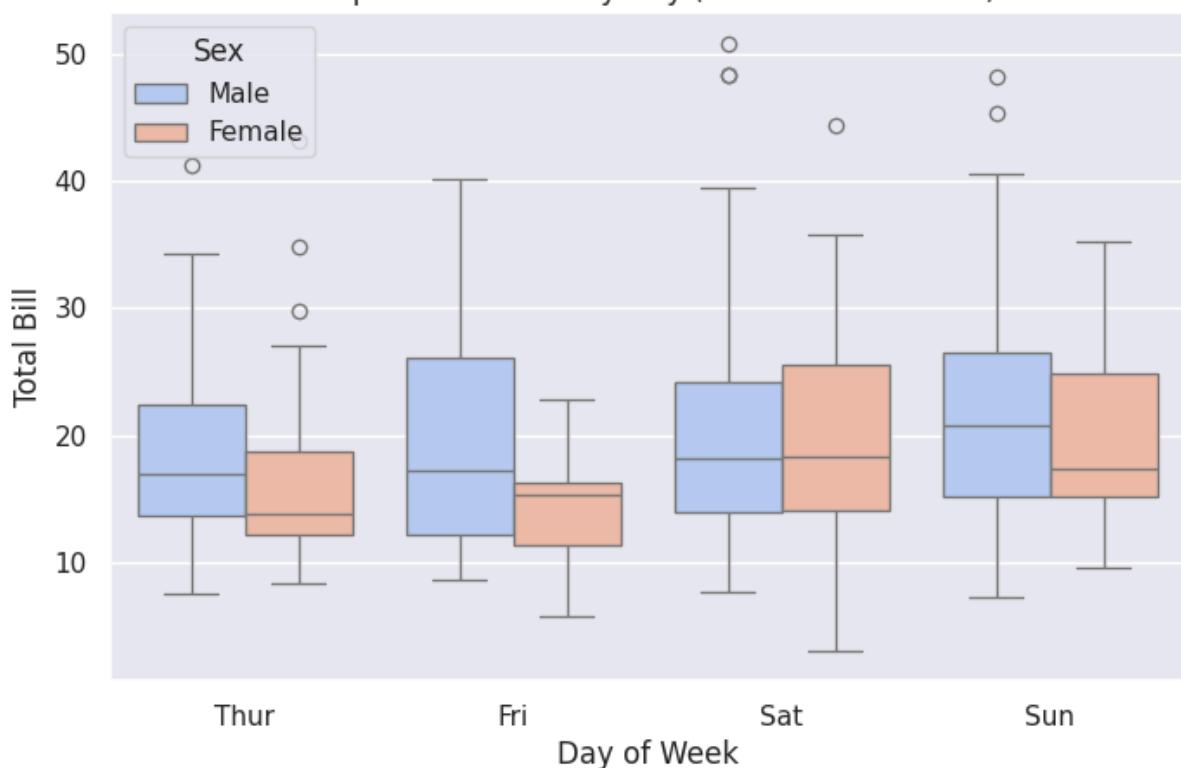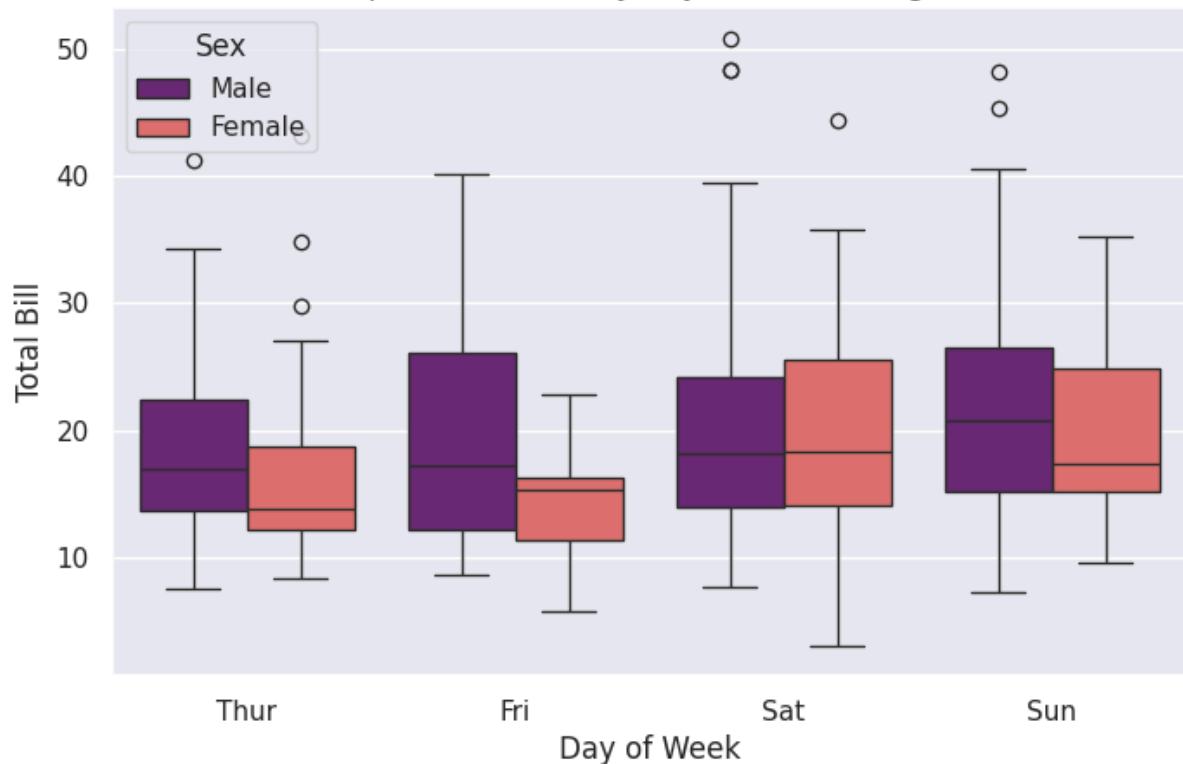
Boxplot of Total Bill by Day (Palette = viridis)

Boxplot of Total Bill by Day (Palette = coolwarm)

Boxplot of Total Bill by Day (Palette = magma)

Distribution of Total Bill (Tips Dataset)

Flights Heatmap (Passengers per Month and Year)

Regression: Total Bill vs Tip by Sex