## Lesson 4: Assignment Details

Your goal for this assignment is use one dimensional and time series analysis techniques from Lesson 4 to create visualizations that provide insight into patterns over time in more than 7 million flights from about a dozen US airline carriers. Using visualization techniques learned in this lesson, please design a visualization to answer each question (4 visualizations total), and provide a brief written analysis to summarize your findings and explain your design rationale.

1.) Examine the quarterly distribution of flights for 2010 and 2011. How do the quarterly trends in the number of flights for 2010 and 2011 compare? Design a visualization that clearly conveys the answer to this question and provide a brief written analysis to: (a) summarize the results and (b) explain your design rationale (e.g., explain why the graph type you selected was effective in light of preattentive attributes).

⠿ Columns	⊟ QUARTER(Flight D..
☰ Rows	CNT(Flight Date)

Quarterly Distribution of Flights for 2010 & 2011
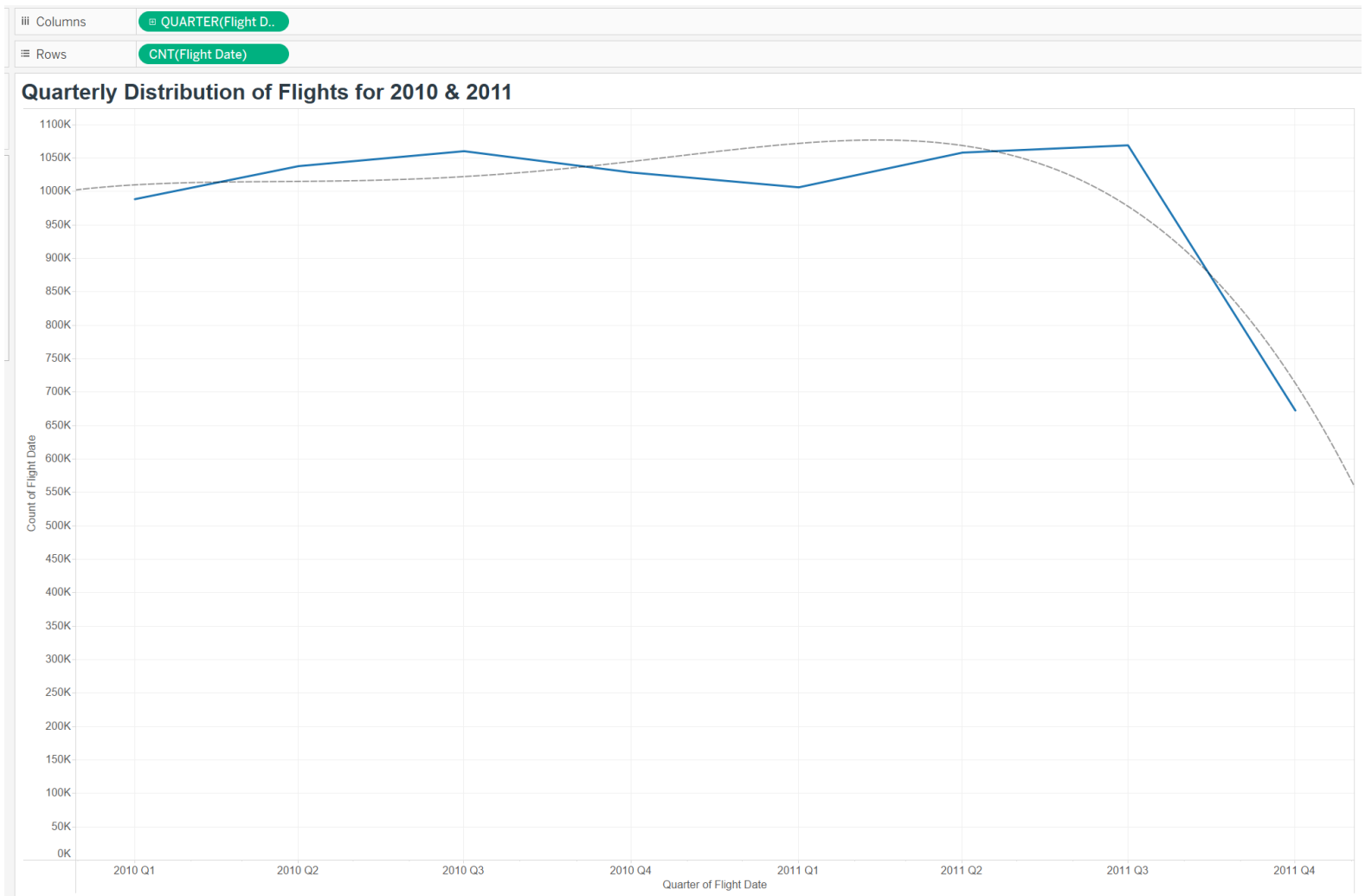
The quarterly trends in data, showed in the year of 2010 each quarter had a more constant frequency of flights. Mainly because in each quarter, the total frequency of flights averaged around 1,000,000 flights scheduled. However, in the year of 2011, the frequency of flights was a little more unstable compared to 2010. Since, from Quarter 1 to Quarter 3 there was a gradual increase of flights scheduled, from around 1,000,000 to 1,070,000 flights. Then from Quarter 3 to Quarter 4 there was a sharp decrease in flights scheduled, from 1,070,000 to 670,000 flights.

In terms of graph selection, I chose a line graph for this problem because line graphs are great for visualizing how data trends over a certain duration of time. And for this problem, we needed to find quarterly trends in data from 2010 to 2011. Which ultimately meant we were dealing with a time series question. We were not deal with a distribution of data question that box plots and histograms are better at solving. And I kept color hues and saturations to a minimum, to not deter away the message the line graph was trying to give to its audience.

2.) Next, explore the flight data more deeply by designing a visualization that will allow you to compare the number of monthly flights in 2010 and 2011. How do the monthly trends for 2010 and 2011 compare? Design a visualization that clearly conveys the answer to this question and provide a brief written analysis to: (a) summarize the results and (b) explain your design rationale.
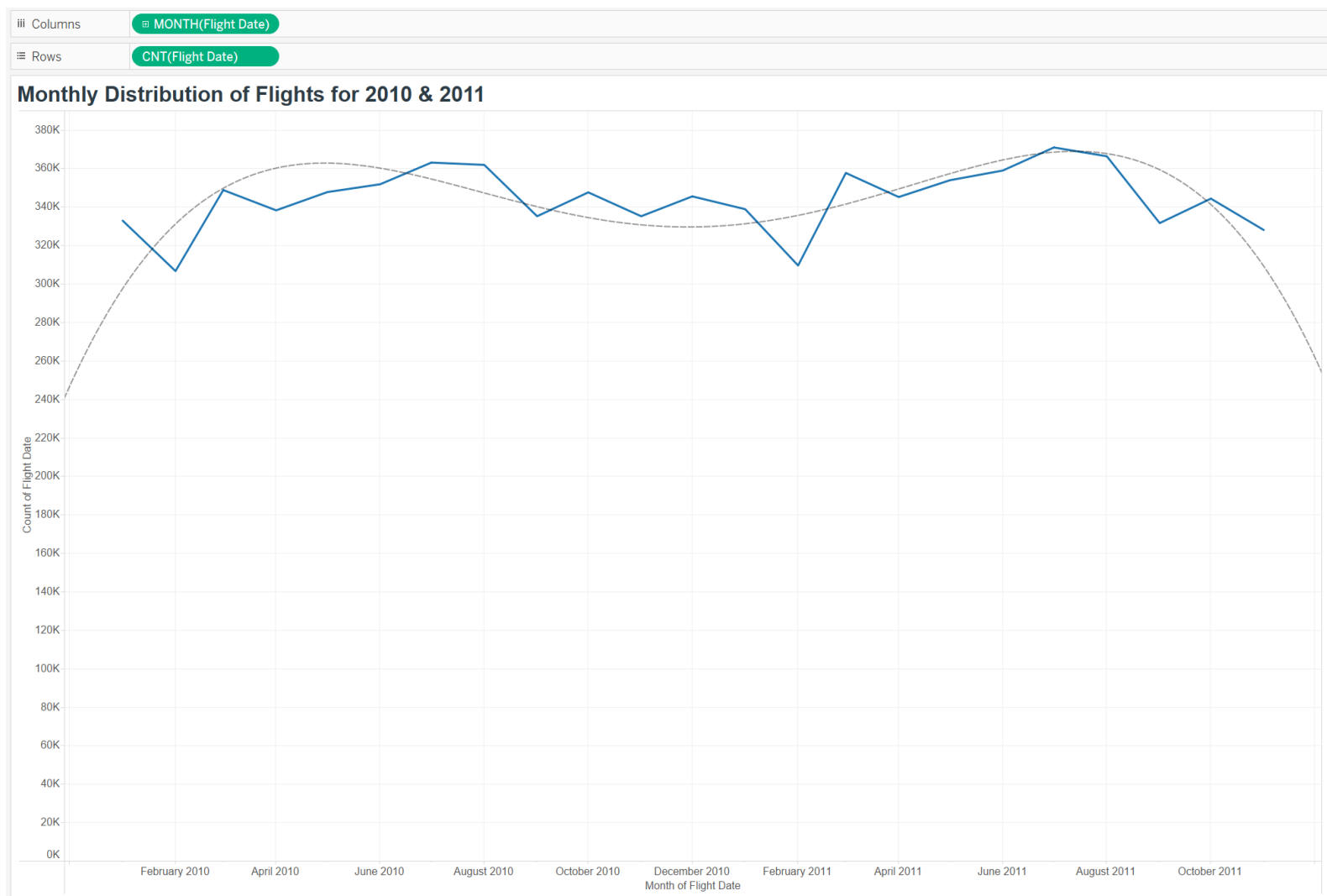

When compared to quarterly trends in flights, the monthly frequency in flights were more stable and constant in the years from 2010 to 2011. Since, on average most airlines had around 340,000 flights each month of the year from 2010 to 2011. There were some flight dips and rises throughout the months, but they eventually averaged out. The biggest increase and decrease in flights between months happened in the year 2011. From February 2011 to April

2011 the frequency of flights increased around 48,000 flights. And from August 2011 to October 2011 the frequency of flights decreased around 35,000 flights.

In terms of graph selection, I chose a line graph again because we were still dealing with a time series question. We were still trying to compare monthly flight trends from the years of 2010 to 2011. Which meant, we were not dealing with a distribution of data question that box plots and histograms are better at solving. I kept color hues and saturation to a minimum to not deter away the message the line graph was trying to give to its audience, as well. However, the use of the polynomial trend line and preattentive attributes such as form (length, width, curvature, texture) and color helped show how monthly frequency of flights changed throughout the years from 2010 to 2011.

3.) You are interested in identifying days that are particularly light in terms of travel each year. Are there particular holidays that seem to experience substantially lower numbers of flights? Design a visualization that clearly conveys the answer to this question and provide a brief written analysis to: (a) summarize the results and (b) explain your design rationale.
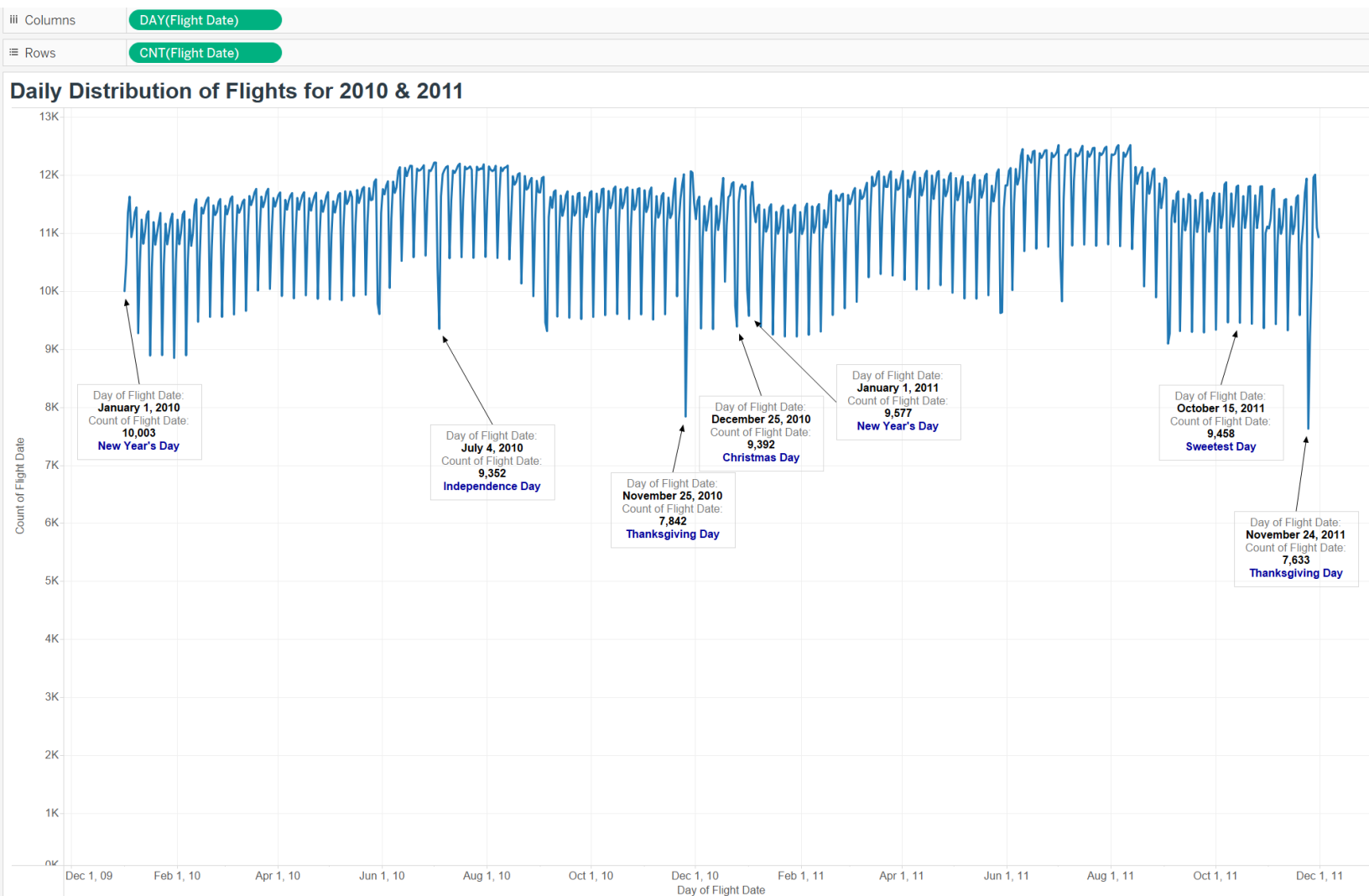
When visualization the flight data by daily aggregation, we could see there were several days that experience substantially lower flight totals. And several of these days fell on US holidays. These holidays included New Year's Day, Independence Day, Thanksgiving Day, Christmas Day, and Sweetest Day. On New Year's Day in the year 2010 there were only 10,003 flights and 9,577 flights in the year 2011. Then on July 4th in the year 2010 there were only 9,352 flights. On Thanksgiving Day in the year 2010 there were only 7,842 flights and 7,633 flights in the year 2011. On Christmas Day in the year 2010 there were only 9,392 flights. And on October 15, 2011, also known as Sweetest Day, there were only 9,458 flights scheduled. Out of all the holidays listed above Thanksgiving Day in the year 2011 had the least number of flights scheduled.

In terms of visualization design, I decided to use a line graph because we still were trying to show trends in flights, which was similar to the previous questions listed above. However now we needed to aggregate the data by days instead of months and quarters. Since, holidays are normally the duration of a single day and not a week or month. And ultimately, we were still dealing with a time series question. However, I used slightly different preattentive attributes to enhance the visualization display of the overall message needed for this question. These preattentive attributes included color, form, and position. I used annotations in the form of rectangular boxes to show each of the holidays that experienced low flight totals and positioned them close to their respective data point on the line graph with arrows. I also highlighted each holiday name in each annotation with a dark intensity of the color purple. This way these data points could stand out from the rest of the data points displayed on the graph.

4.) Next, you are interested in examining the flight cancellation data per airline. Which airlines had the highest number of cancelled flights in 2010 and in 2011? Design a visualization that clearly conveys the answer to this question and provide a brief written analysis to: (a) summarize the results and (b) explain your design rationale.

Based on the bar chart listed below, we were able to discover that Southwest Airlines had the highest number of flight cancellations in 2010 and 2011. Southwest airlines had considerably more flight cancellations compared to the other airlines listed in this dataset with 2,181,729 flights cancelled. The second most flight cancelled were from Delta Airlines with around 1,400,000 flights cancelled.

The reason why I chose a bar chart for this visualization was because I needed to compare changes in flight cancellations between groups; these groups happen to be different airline carriers. And the flight cancelled, needed to be measured over a long duration of time, two years. Which is why I chose a bar chart instead of a line graph, for this question. We also saw with our daily aggregated line graph, as time series duration increases, data points seem to get clustered together. Which makes it slightly harder to depict trends in data. Bar charts, on the other hand, handle large ranges of data points better than line graphs in this regard. I still was able to use annotations and ranges of color intensity to show differences in flight cancellations among airlines, though.