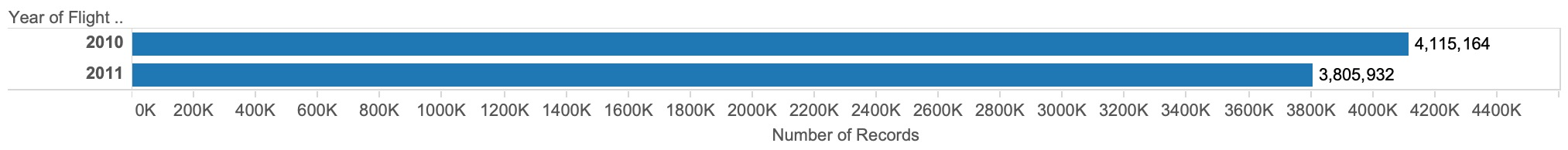
**JEREMY PETERSON DUTOYA**

**PR. EMRE YETGIN – DIRECTOR, CENTER FOR BUSINESS ANALYTICS**

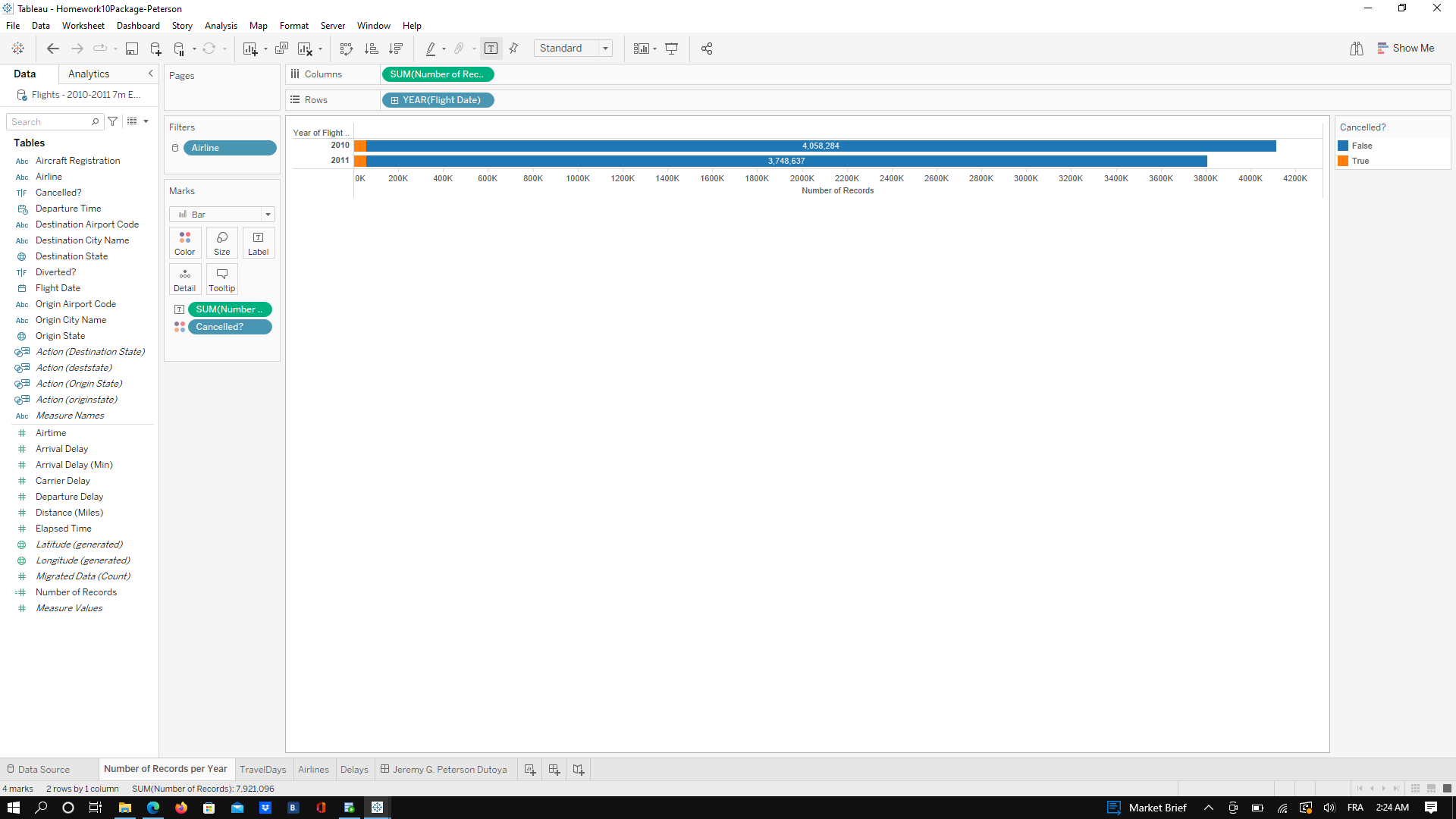
**PMBA-8355-OLB VISUAL ANALYTICS**

* *Finding Answers in the Data:*

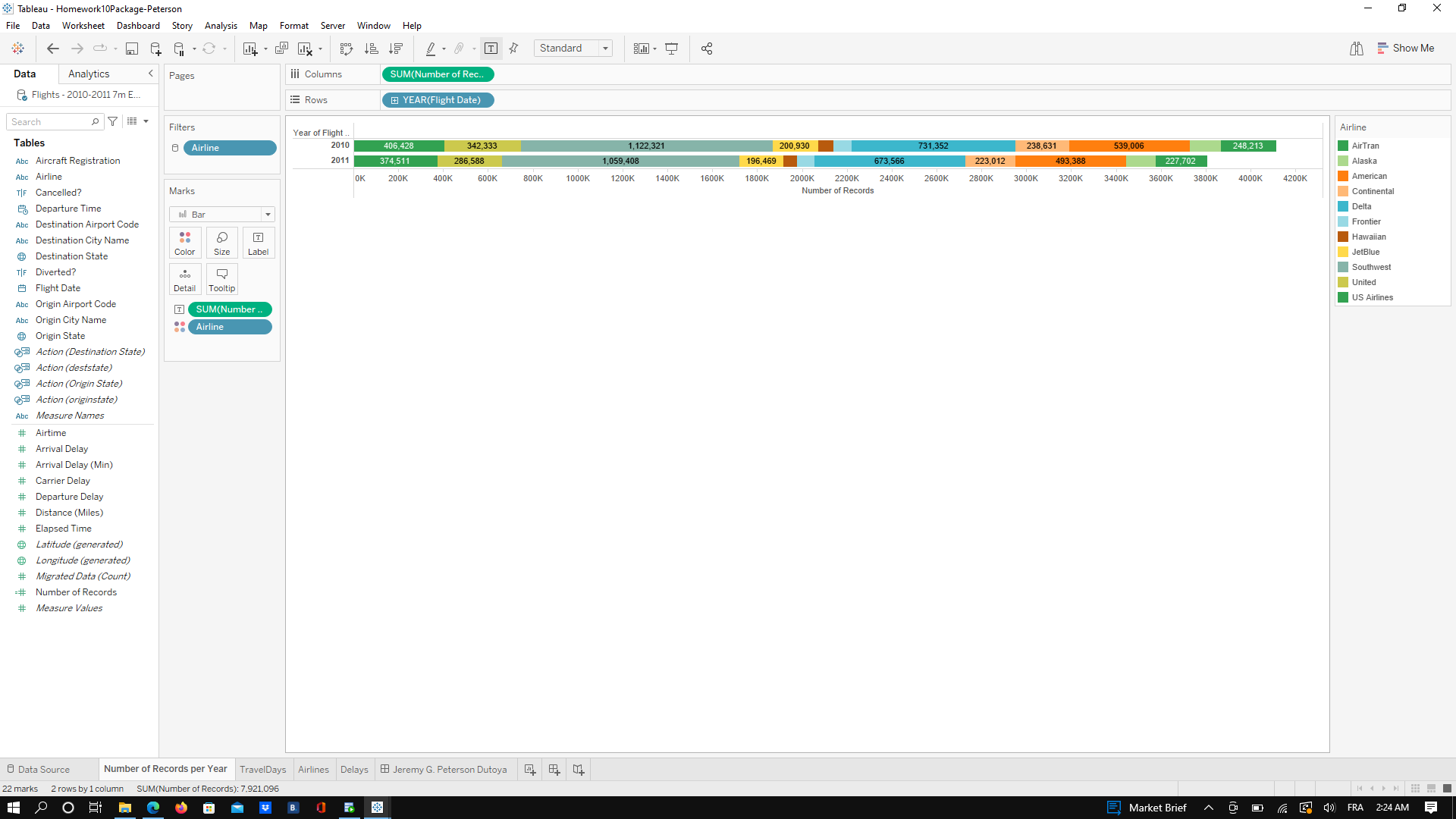
*Before I start with the answers I found for my homework, I would like to share how much I enjoyed working on this assignment. I found it very interesting because it asks us to use all the methods and best practices we have learned since the beginning of this class to answer business-related questions that can be useful without you showing us and guiding us to find the answers. I started this assignment by telling myself that I was going to answer one question a day but I was taken by a craze, as in a puzzle game or a detective movie, and here I am at 3 am still working on it and playing with the data. That is how much I like it. Thank you for this kind of homework Professor Yetgin.*

1. A quick visualization over time shows a drop of more than 300,000 flights from 2010 to 2011:

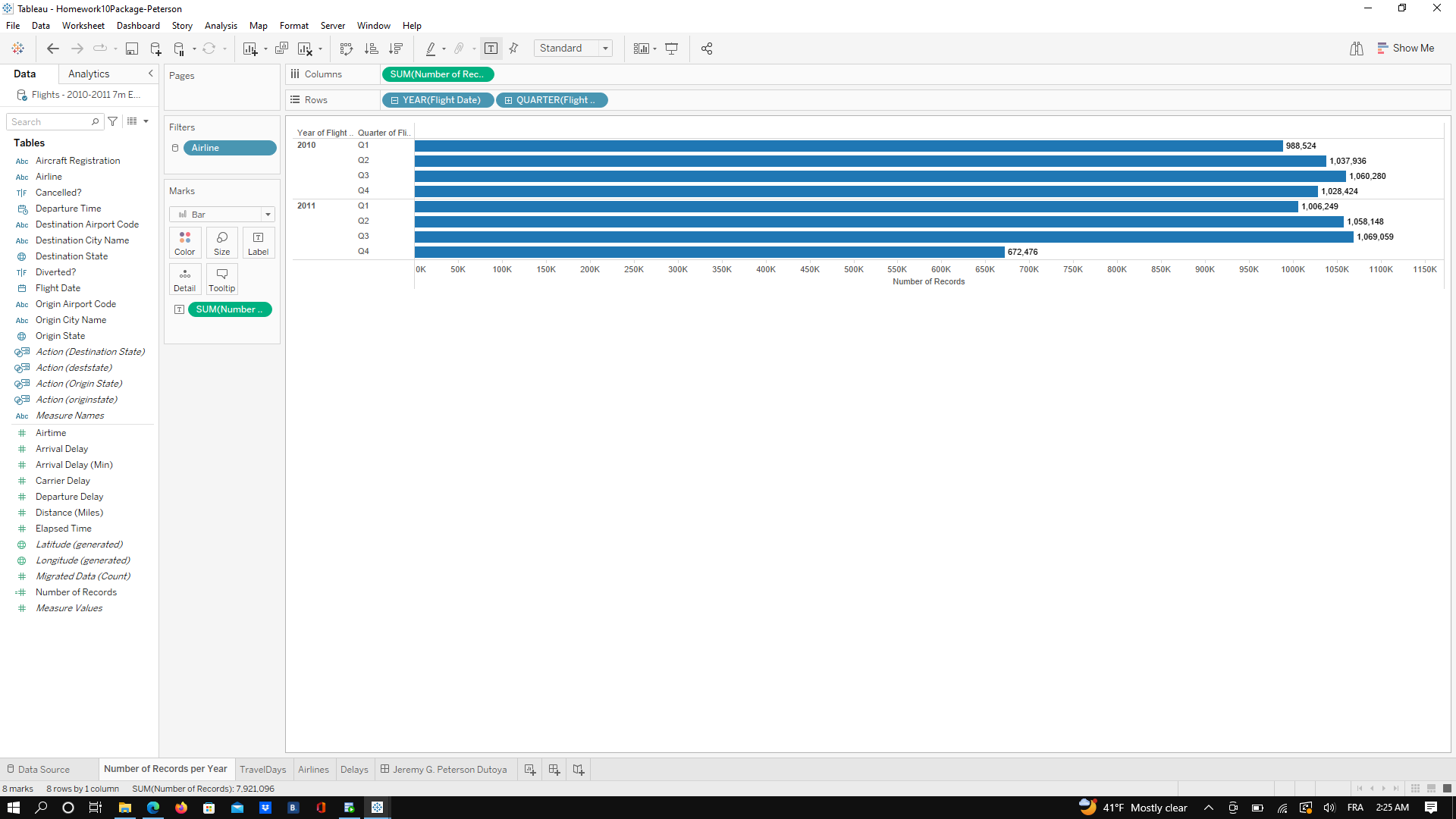
*You were certain that flight activity was not really different between 2010 and 2011, even though this visualization seems to indicate otherwise.* ***Investigate flights over time in more detail and explain why these yearly numbers indicate such a large drop.***

For this visualization, I first thought: “Oh, Maybe there have been more cancellations in the year 2011 causing the drop.” I then created the following visualization to see if I was right: 

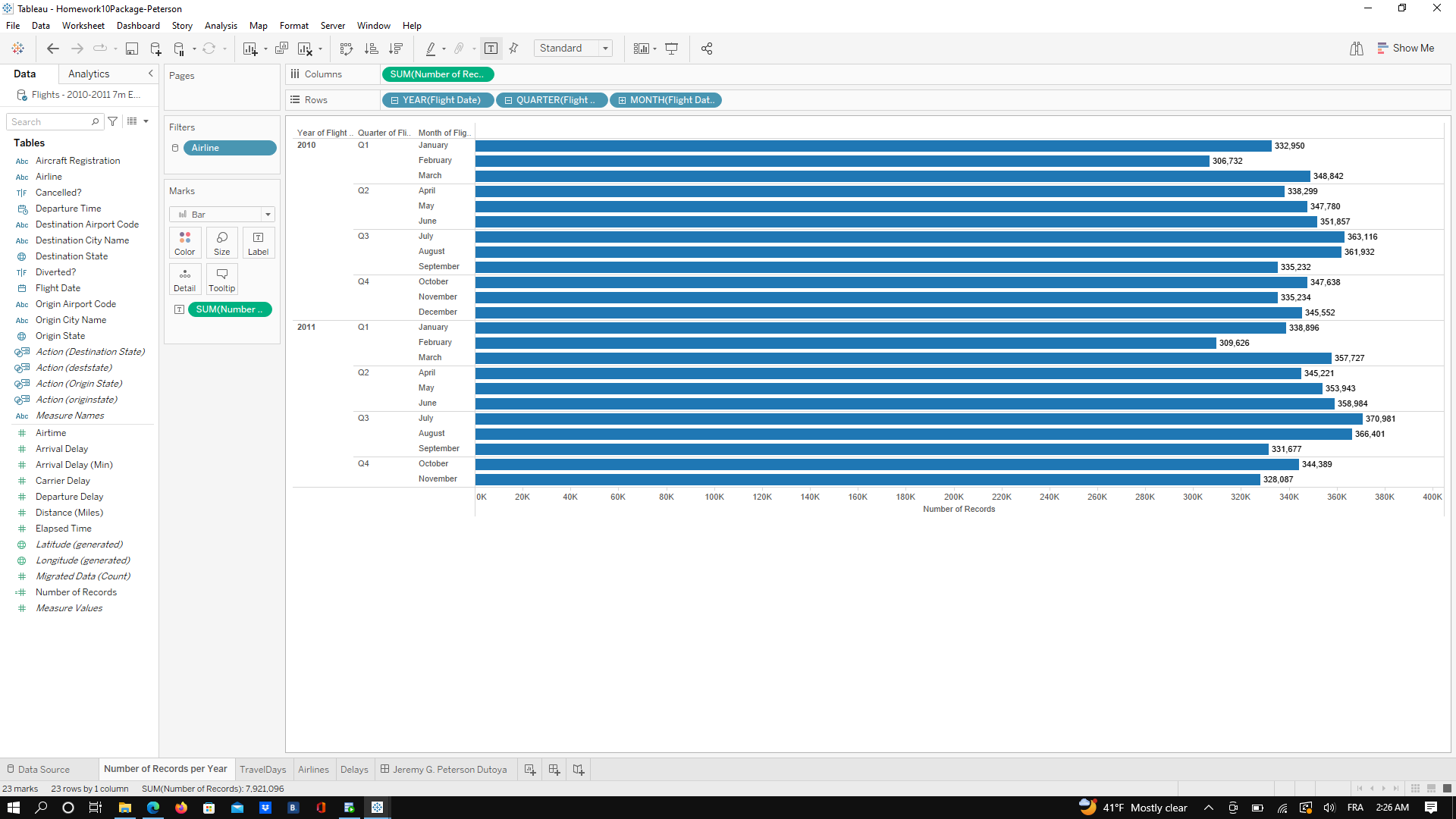
No, I was not. I was very much wrong because as we can see the number of cancellations in 2010 and 2011 is about the same. I was not discouraged, and kept looking for answers! I then thought: “Maybe one or more airline companies went bankrupt and stopped providing flights in 2011 causing the drop in the number of flights.” So, I added Airline to the Color Mark and I created the following visualization:



Once again, I was wrong. The number of Airlines is the same in 2010 and 2011 and we have the exact same list of airlines and they all offered approximately the same number of flights in these 2 years. I was not discouraged but even more excited to find what was hiding in the data. I started playing with the Year, Quarter, and Month and I saw something interesting appearing before my eyes!

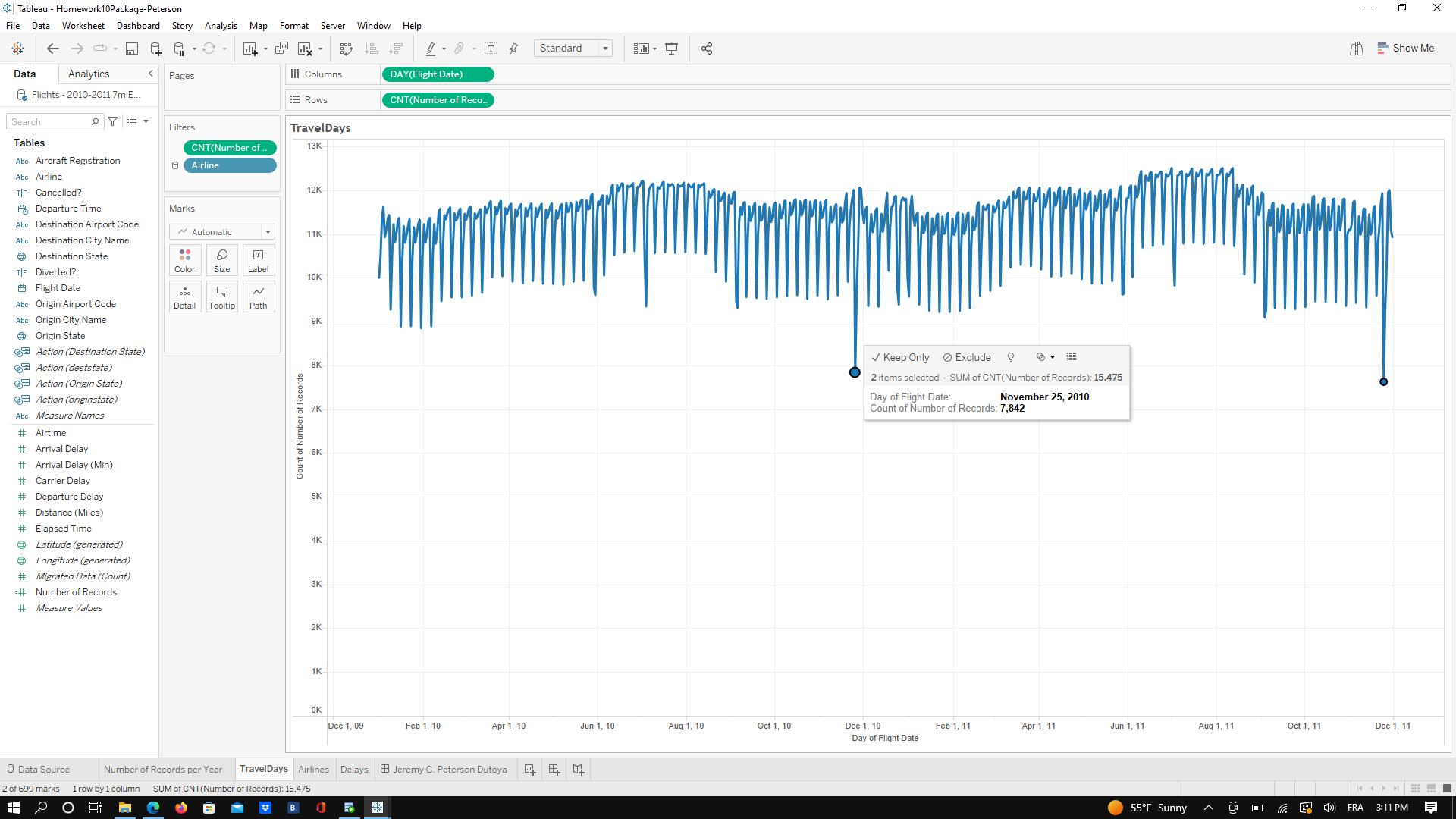


With this visualization, we can see that the drop occur in the Q4 of 2011 and that every other Quarter of 2011 is a little superior in terms of records than the quarter in 2010. So what is happening in Quarter 4 of 2011 that is causing this huge drop?



Now we are able to see that the reason for this 300,000 drop in terms of the number of Records is that we have **no records for the Month of December 2011**. This can happen if they collected the data before December, or something happened during the Data Cleaning process or several other reasons. But, in order to compare, in the month of December 2010, the number of records is about 345,000 so the fact that the month of December is missing really explains this drop from 2010 to 2011.

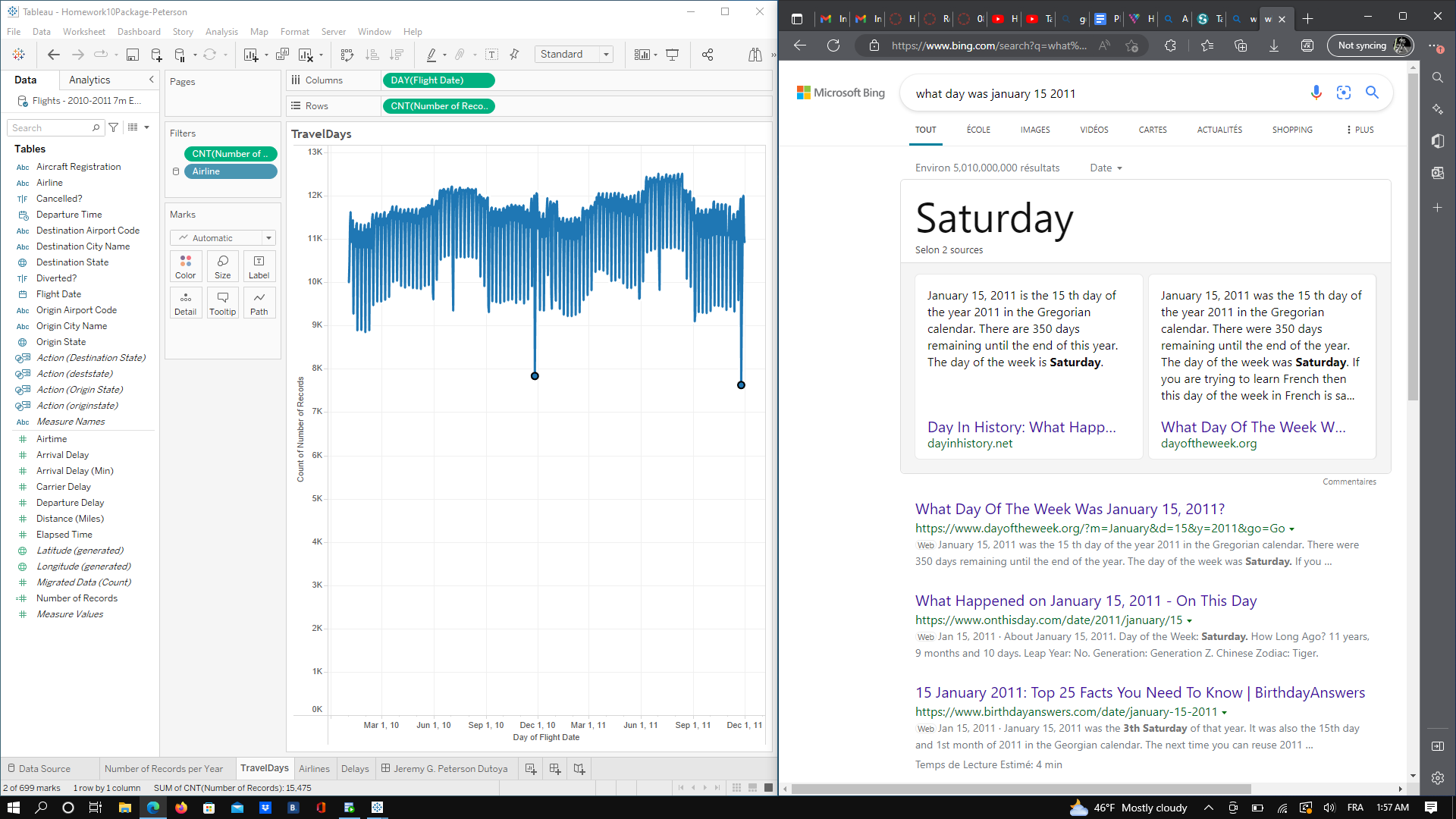
2. You want to identify the least traveled days for each year. **Are there particular holidays that seem to experience a substantially lower number of flights? Which one(s)? Why do you think so?**

****

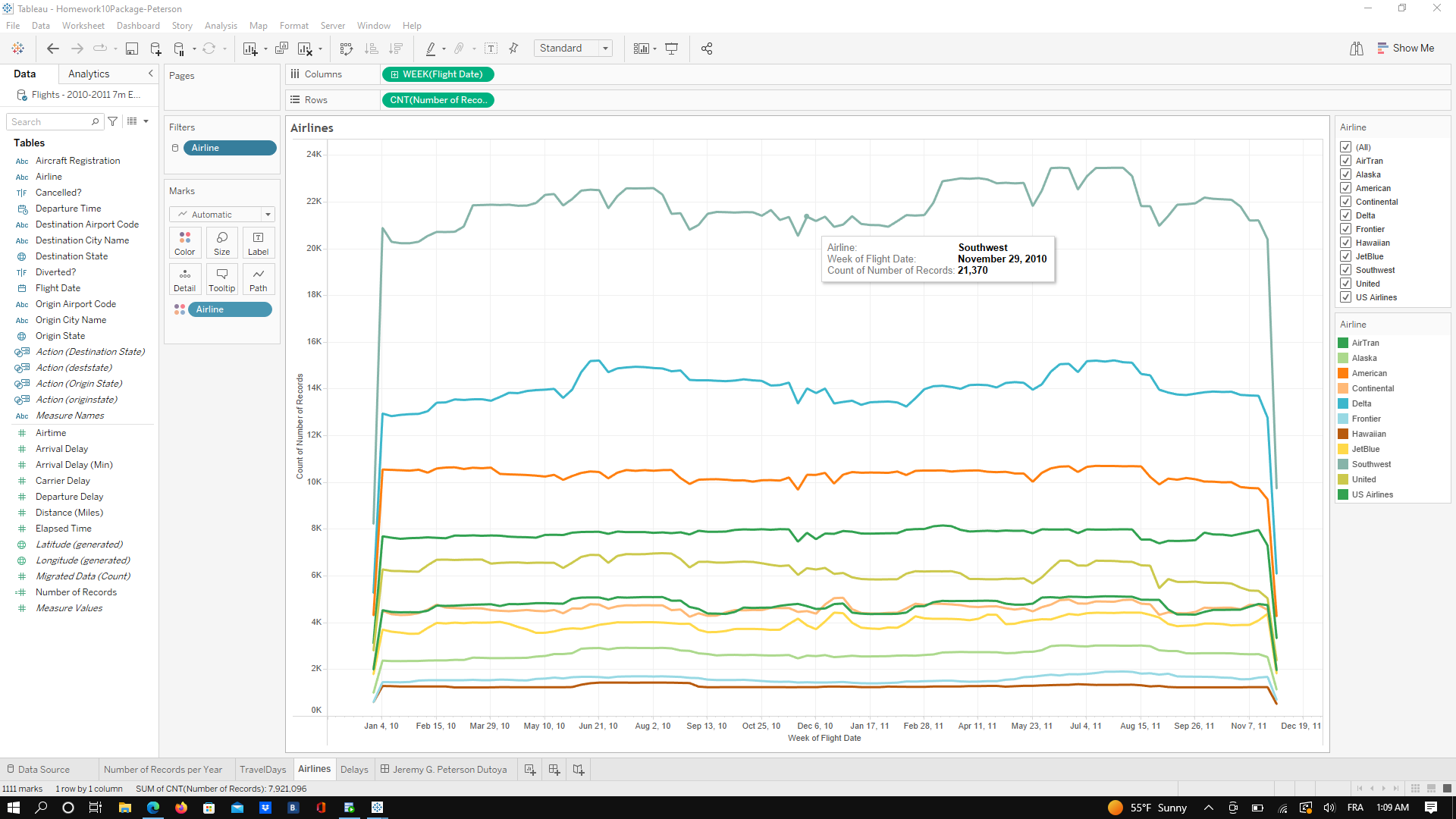
Thanks to this visualization we can see that there are several days in the year when the number of flights is experiencing a drop. In particular, we can see that the two days that are experiencing the most significant drop are November, 25 of 2010, and November, 24 of 2011: This is **Thanksgiving.** This is easily understandable why there is such a drop in flights on thanksgiving because many Americans consider Thanksgiving to be a very meaningful holiday, and for some people, it has more meaning than Christmas Day, Hanukkah, or Ramadan. On Thanksgiving, people stay with their families to eat, share a beautiful meal and be grateful for the people that share their lives with them. For this reason, you don’t want to be on a flight on the day of thanksgiving but with your loved ones to give thanks and count the blessings in your life.

Other Holidays that experience a drop in this dataset are:

* the **4th of July,**
* September 5th which is **Labor Day**,
* December 25th which is **Christmas**,
* May 30th of 2010, and May 28th of 2011 which were both one day away from **Memorial Day,**
* **Saturdays in general** (I looked almost every day that were experiencing a drop in the number of flights and they almost all ended being a Saturday: January 9, 2010 - January 16, 2010, January 23, 2010 - January 30, 2010 - February 6th, 2010 and so on until January 15th, 2011 all these days represent a lower number of flight and they are all Saturdays)



3. You want to know which airline operates the greatest number of flights. **Is there an airline that operated more flights than any other single airline for each week in this dataset? Which one?**



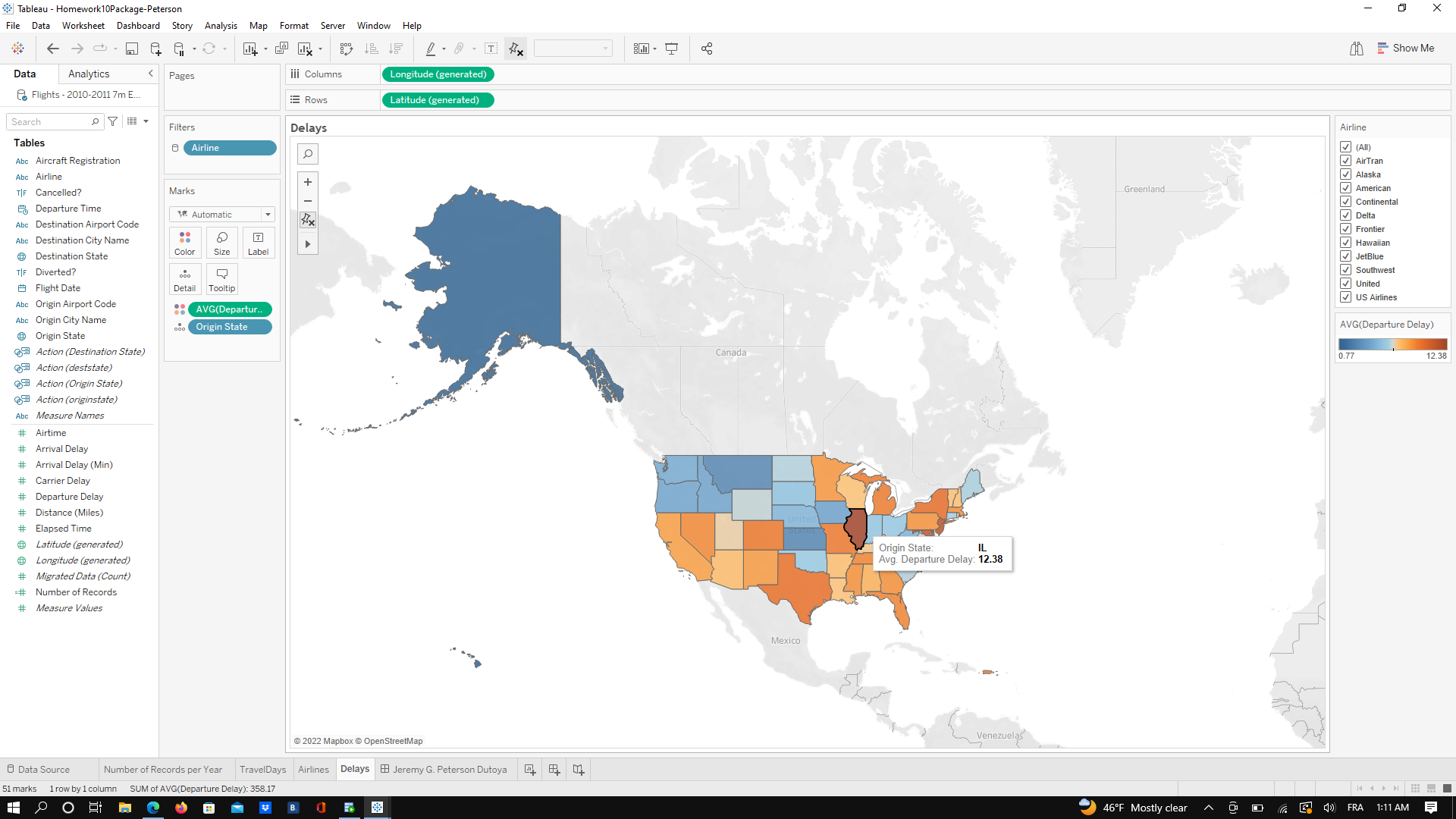
As it is represented and clearly made visible in this visualization, **Southwest** is the Airline that operates the most flights each week than any other airline.

In order to compare:

* It operated 1.8x more flights than Delta which is second in the Ranking.
* It operated 2x more flights than American which is third in the Ranking.
* It operated 3x more flights than US Airlines which is 4th in the Ranking.
* It operated 4x more flights than United which is 5th in the Ranking.
* It operated 5x more flights than Airtran and Continental which are 6t in the Ranking.
* And so on until Hawaiian Airlines, last in the ranking, which operated 19x fewer flights than Southwest between 2010 and 2011

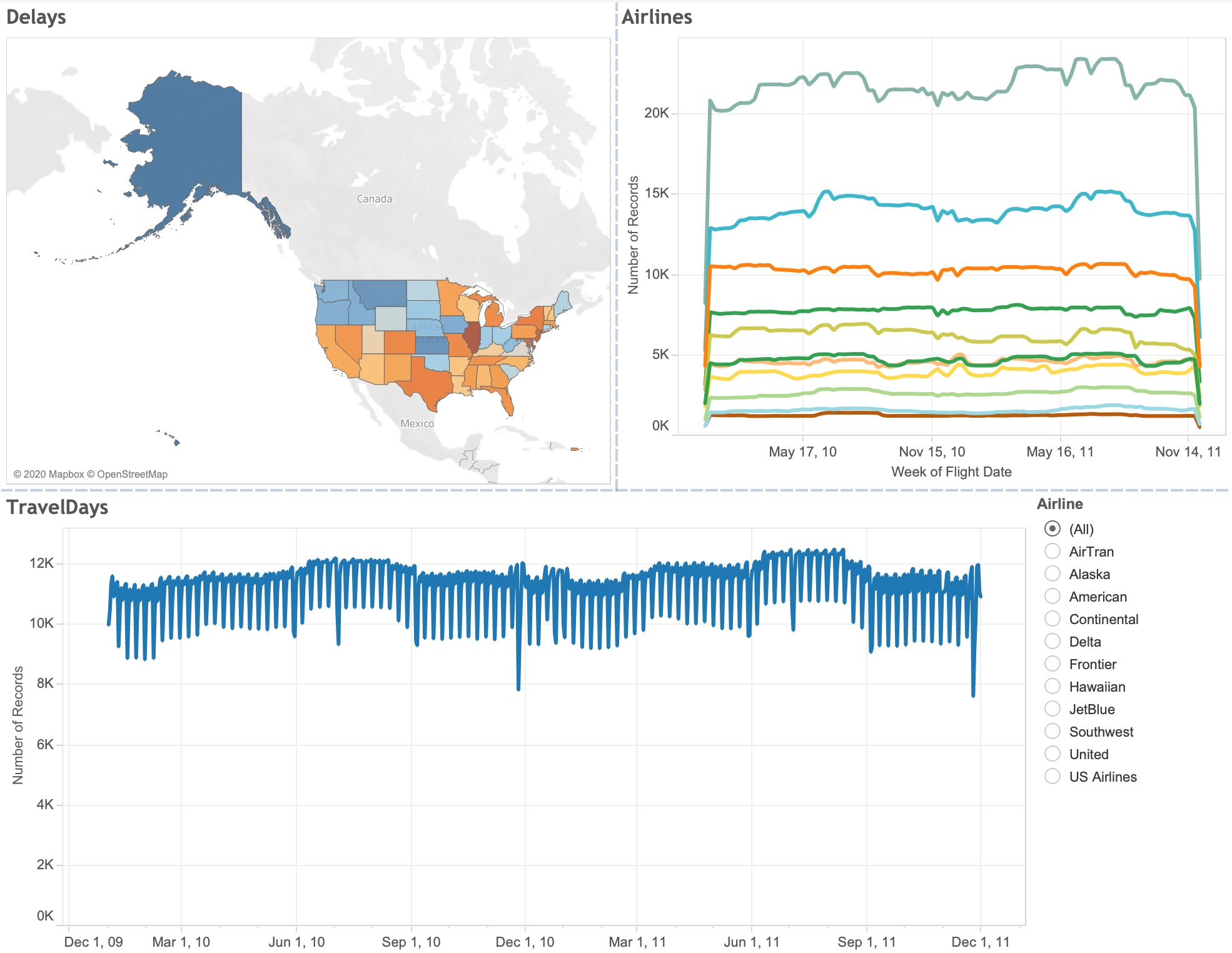
4. You want to know which states tend to experience the longest departure delays. **Create a map that visualizes the average departure delay that you can expect in each origin state. Which state(s) experience the worst delays?**

* + - Allow the user to filter by airline.
    - Colors should range from dark blue for states with short departure delays to dark orange for states with long departure delays.



The State with the most average departure delay is **Illinois (IL) with a 12.38** minutes delay on average.

Organize the worksheets you created into the **dashboard** depicted below.

* Try to recreate this dashboard exactly
* Note that selecting an airline on the filter should update all three visualizations  
  

My Dashboard: (the trick is to use the “blank” dashboard object in order to place the floating “Airline” filter and have the exact same dashboard as you required)

