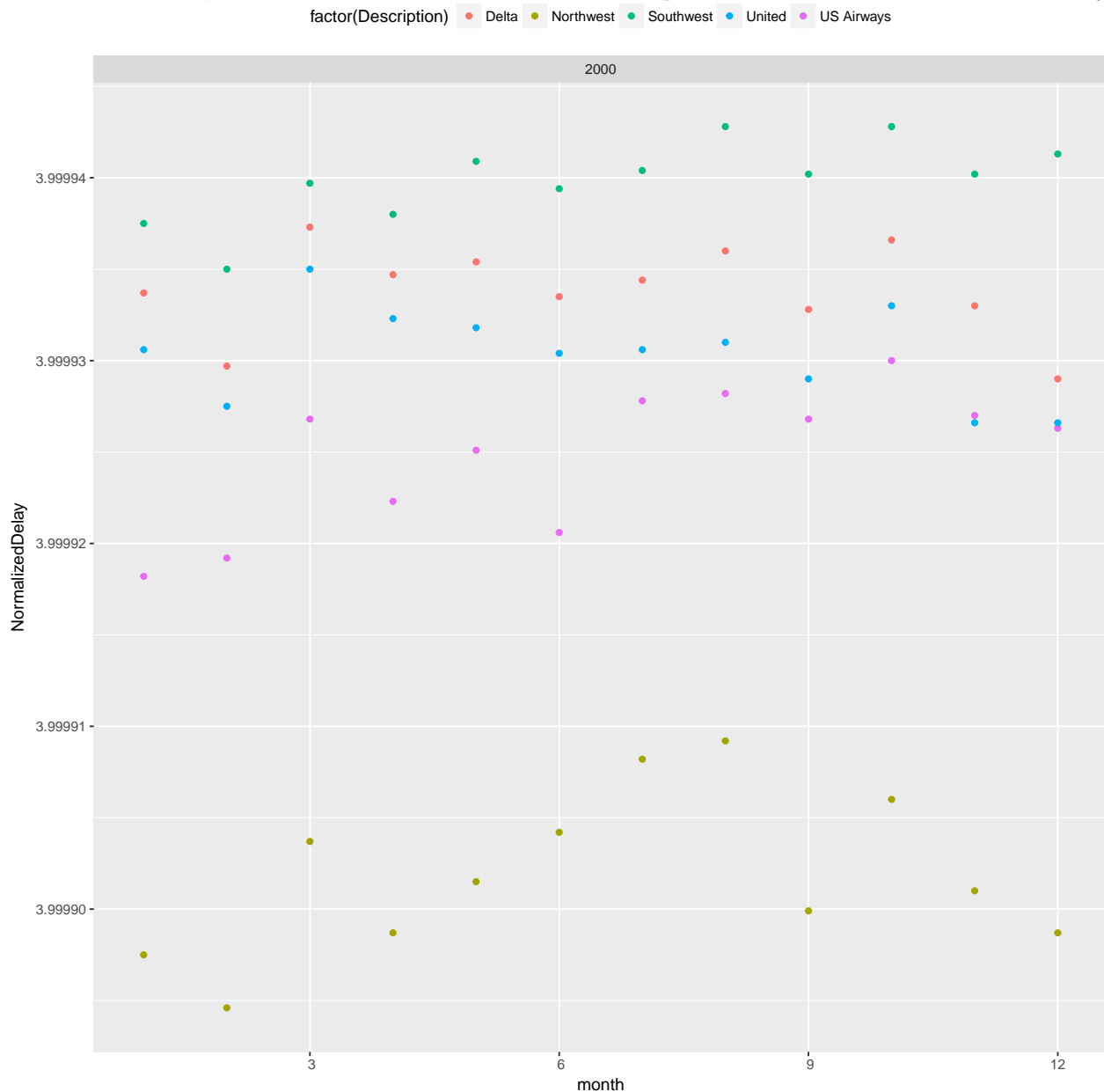
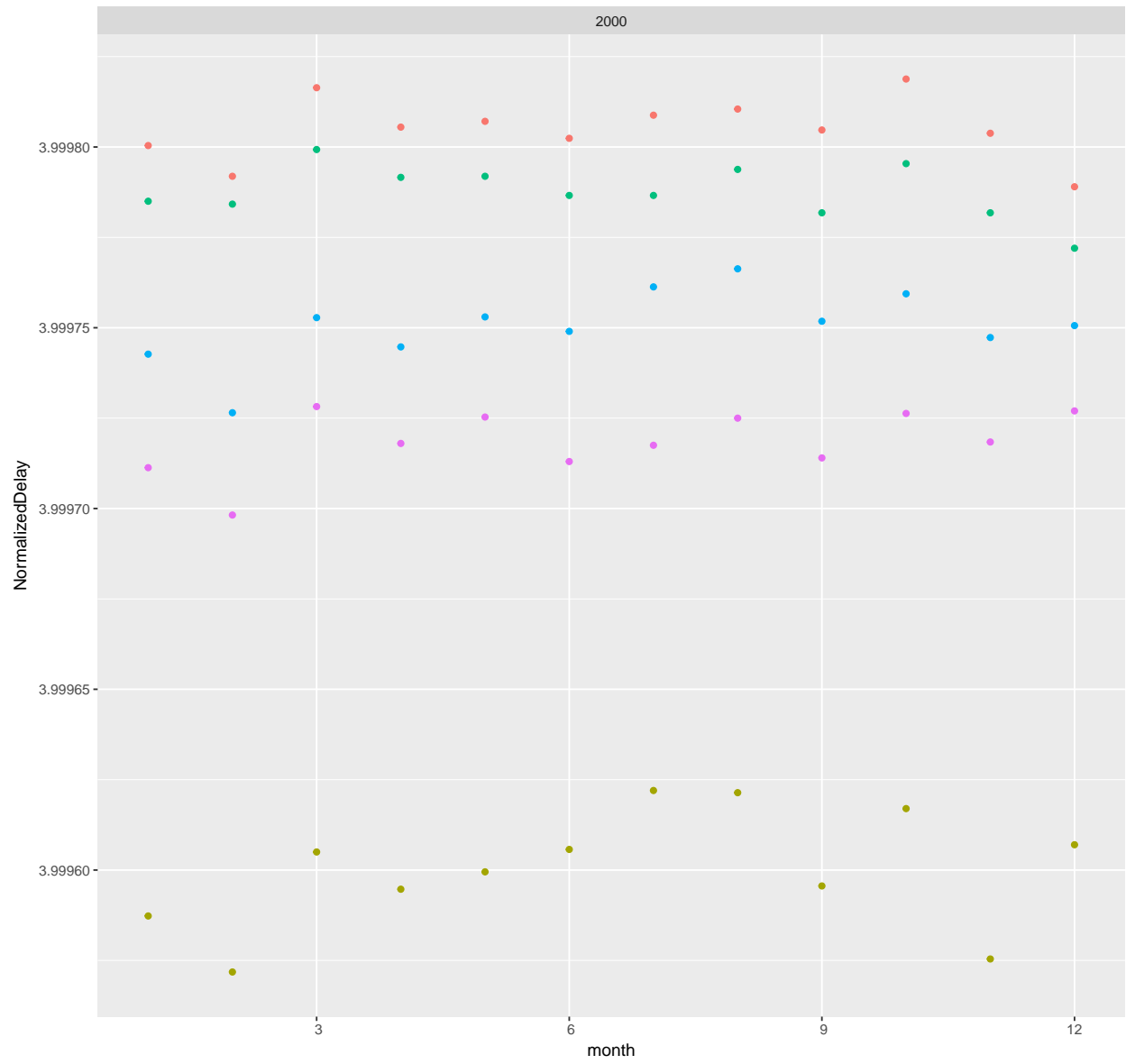


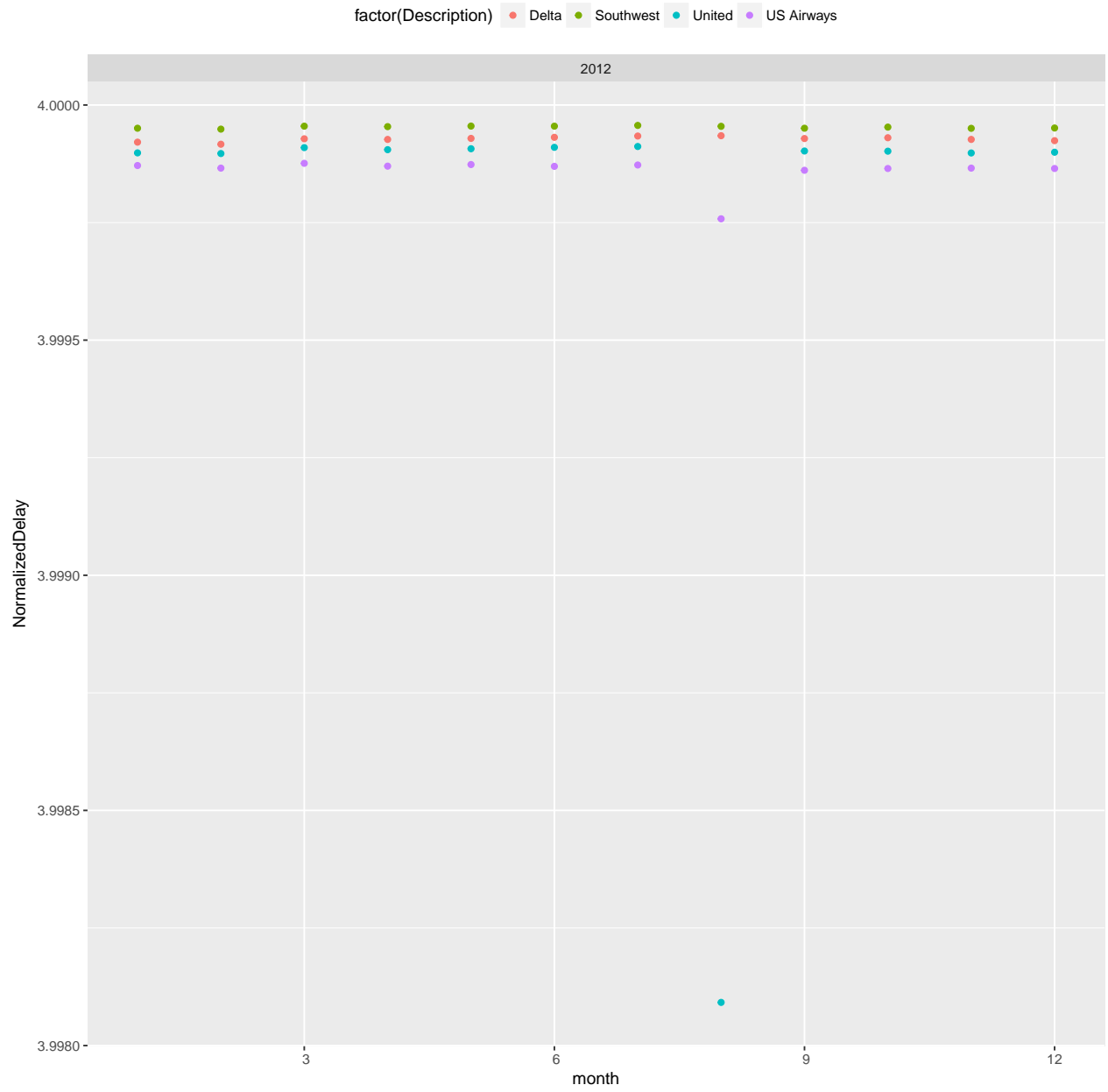
Program Design

The goal of this assignment was to provide visualization mean delay of the five most active airlines and for the five most active airports in the country. I have used 1 Map-Reduce job that does the initial cleaning of the airline dataset and then gets the normalized delay (with respect to the scheduled length of each flight) and mean delay for each month for a particular year. The output is a results.csv file placed in the `output` folder that has `airportId` and year or `airlineId` and year as key and value as `NormalizedDelay` for the year and for each month separated by "," starting from Jan. ##### Plot 1: Most Active Airlines in the period 2011-2015 with their Mean Delay

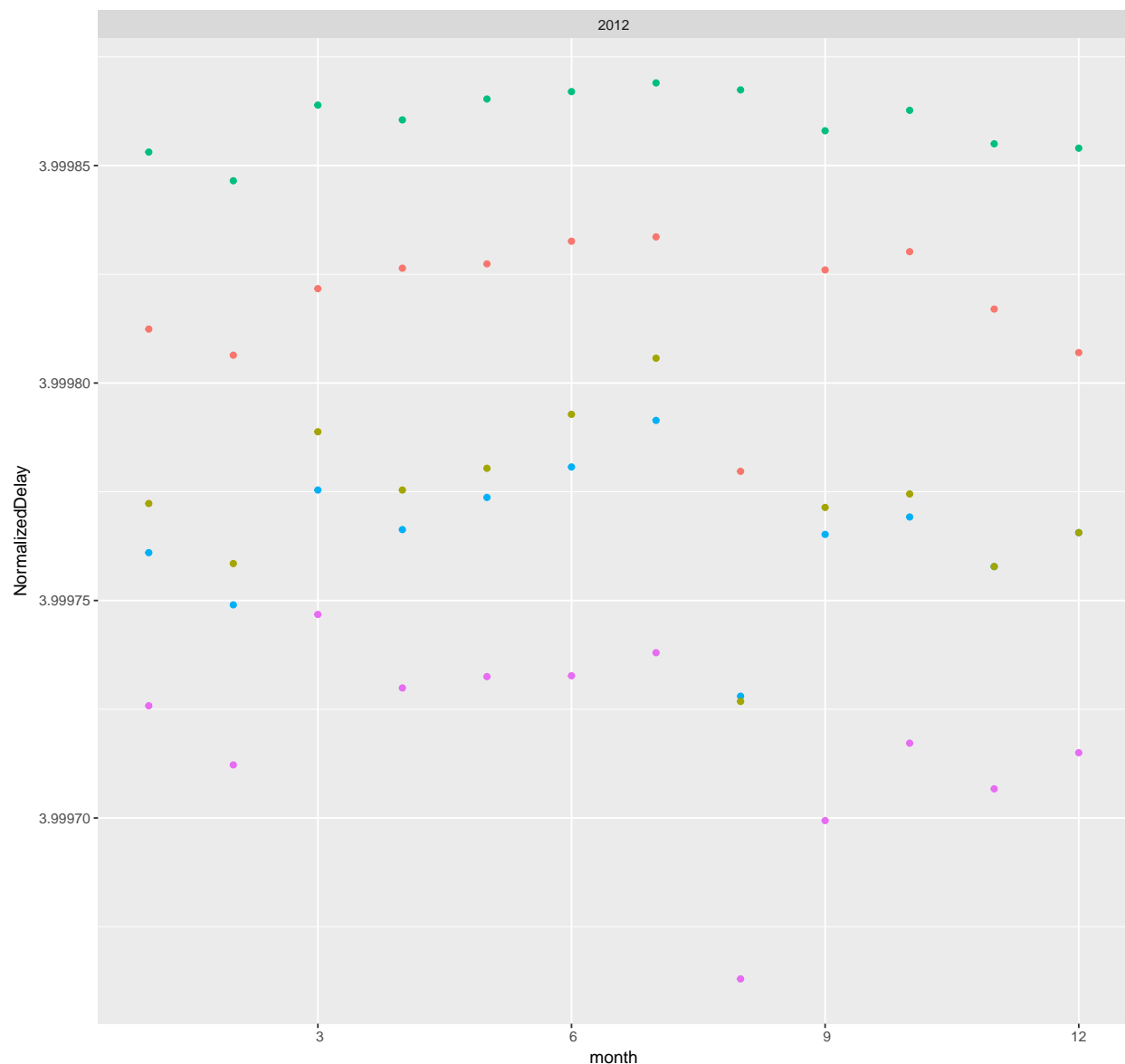


scription) ● Chicago O'Hare International ● Denver International ● Hartsfield-Jackson Atlanta International ● Los Angeles, CA: Los Angeles International ● Phoenix





scription) ● Chicago O'Hare International ● Denver International ● Hartsfield–Jackson Atlanta International ● Los Angeles, CA: Los Angeles International ● Phoenix



The Plots 1-? represent the Normalized delay for the Top 5 airlines. The available data is from Dec 1987 to Oct 2015. For the period 1988 to 1990 US Airways topped the list for delays, however towards the end of 1990 Southwest and Delta were more delayed. Thereafter, for a majority period from 1991 to 1998 Delta was the most delayed flight, closely followed by Southwest and US airways. It is interesting to note that **Northwest Airlines** was the least delayed among these 5 top airlines from 1987 to 2004 but after 2004 it topped the list for delays till the time it was merged with Delta airlines in 2010 Jan due to bankruptcy and closed down operations as Northwest in 2009 Dec. After 1998 Southwest has topped the list for delays among the other 4 airlines. Another point to note is that the airlines are the most delayed during the month of december (this could be because of holiday season).

The Plots 1-? represent the Normalized delay for the Top 5 airports. The available data is from Dec 1987 to Oct 2015. For the period 1988 to 1990 US Airways topped the list for delays, however towards the end of 1990 Southwest and Delta were more delayed. Thereafter, for a majority period from 1991 to 1998 Delta was the most delayed flight, closely followed by Southwest and US airways. It is interesting to note that **Northwest Airlines** was the least delayed among these 5 top airlines from 1987 to 2004 but after 2004 it topped the list

for delays till the time it was merged with Delta airlines in 2010 Jan due to bankruptcy and closed down operations as Northwest in 2009 Dec. After 1998 Southwest has topped the list for delays among the other 4 airlines. Another point to note is that the airlines are the most delayed during the month of december (this could be because of holiday season).