

Investigation of airlines and airports

Roman Melnyk

Tuesday, March 24, 2015

- [Problem 1](#)
- [Problem 2](#)
- [Problem 3](#)

Let's see on "Flights" data set:

```
##   year month day dep_time dep_delay arr_time arr_delay carrier tailnum
## 1 2013     1   1      517         2      830         11      UA   N14228
## 2 2013     1   1      533         4      850         20      UA   N24211
## 3 2013     1   1      542         2      923         33      AA   N619AA
## 4 2013     1   1      544        -1     1004        -18      B6   N804JB
## 5 2013     1   1      554        -6      812        -25      DL   N668DN
## 6 2013     1   1      554        -4      740         12      UA   N39463

##   flight origin dest air_time distance hour minute
## 1   1545    EWR  IAH      227      1400     5      17
## 2   1714    LGA  IAH      227      1416     5      33
## 3   1141    JFK  MIA      160      1089     5      42
## 4    725    JFK  BQN      183      1576     5      44
## 5    461    LGA  ATL      116       762     5      54
## 6   1696    EWR  ORD      150       719     5      54
```

Departure Delay satistic(best and worst airline)

```
##   ID              Name Avarrage.departure.delay
## 1  US      US Airways Inc.          3.744693
## 2  HA  Hawaiian Airlines Inc.          4.900585
## 3  AS    Alaska Airlines Inc.          5.830748
## 4  AA  American Airlines Inc.          8.569130
## 5  DL    Delta Air Lines Inc.          9.223950
## 6  MQ      Envoy Air              10.445381
## 7  UA    United Air Lines Inc.       12.016908
## 8  OO    SkyWest Airlines Inc.       12.586207
## 9  VX      Virgin America          12.756646
## 10 B6      JetBlue Airways          12.967548
## 11 9E      Endeavor Air Inc.        16.439574
## 12 WN    Southwest Airlines Co.     17.661657
## 13 FL AirTran Airways Corporation   18.605984
## 14 YV      Mesa Airlines Inc.       18.898897
## 15 EV    ExpressJet Airlines Inc.   19.838929
## 16 F9      Frontier Airlines Inc.   20.201175
```

Arriving delay statistic

```
##   ID              Name Avarrage.arriving.delay
## 1  AS    Alaska Airlines Inc.       -9.9308886
## 2  HA  Hawaiian Airlines Inc.       -6.9152047
## 3  AA  American Airlines Inc.        0.3642909
## 4  DL    Delta Air Lines Inc.        1.6443409
## 5  VX      Virgin America            1.7644644
## 6  US      US Airways Inc.           2.1295951
## 7  UA    United Air Lines Inc.        3.5580111
## 8  9E      Endeavor Air Inc.         7.3796692
## 9  B6      JetBlue Airways           9.4579733
## 10 WN    Southwest Airlines Co.      9.6491199
## 11 MQ      Envoy Air                 10.7747334
## 12 OO    SkyWest Airlines Inc.       11.9310345
## 13 YV      Mesa Airlines Inc.        15.5569853
## 14 EV    ExpressJet Airlines Inc.    15.7964311
## 15 FL AirTran Airways Corporation    20.1159055
## 16 F9      Frontier Airlines Inc.    21.9207048
```

After analyzing previous two tables we can clearly see best and worst airlines groups. "Frontier Airlines" is the worst airline. As we can see this airline has in average the longest departure and arriving delays. The best are "Alaska Airlines" and "Hawaiian Airlines"

Problem 1

Which airline is the most expeditious for the longest distances?

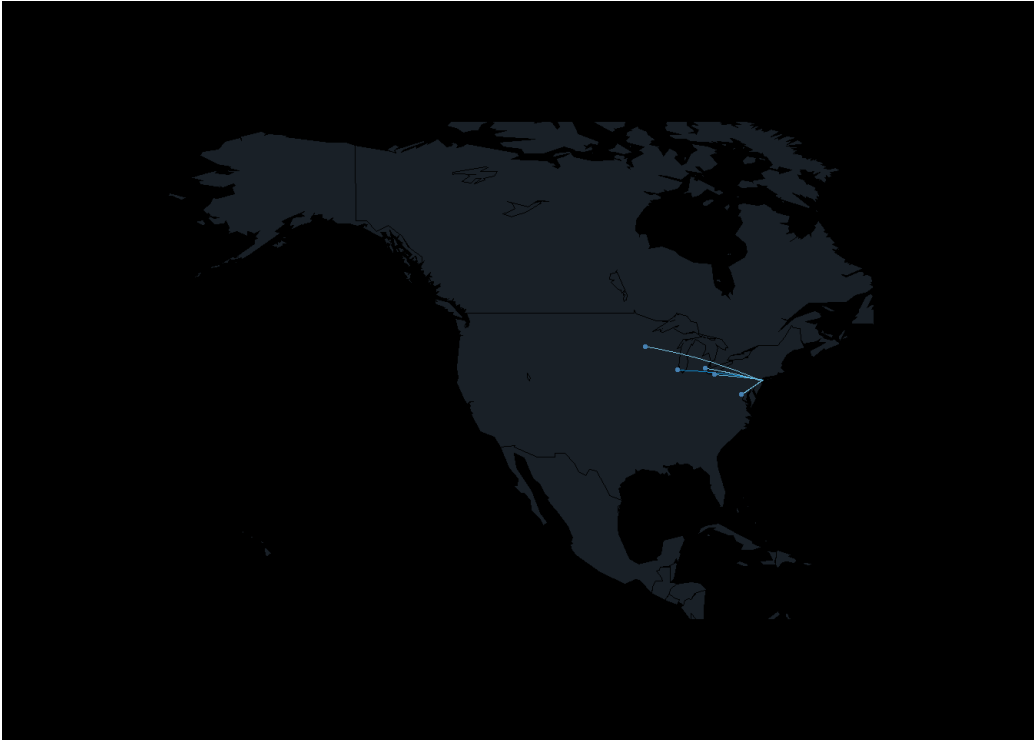
We should find out airline than spends the least time to overcome the longest distance. The problem is for every airline to find the longest route and then exact the minimum time that goes for overcoming it. For example, the longest route for “American airlines” is 3500, but at different times of years spends different time for overcoming it. So, our job is to find the least time. And later we will compare speed for longest routes with general speed for each airline, result will be very interesting.

First, let’s see how looks routes.

Southwest Airlines



SkyWest Airlines

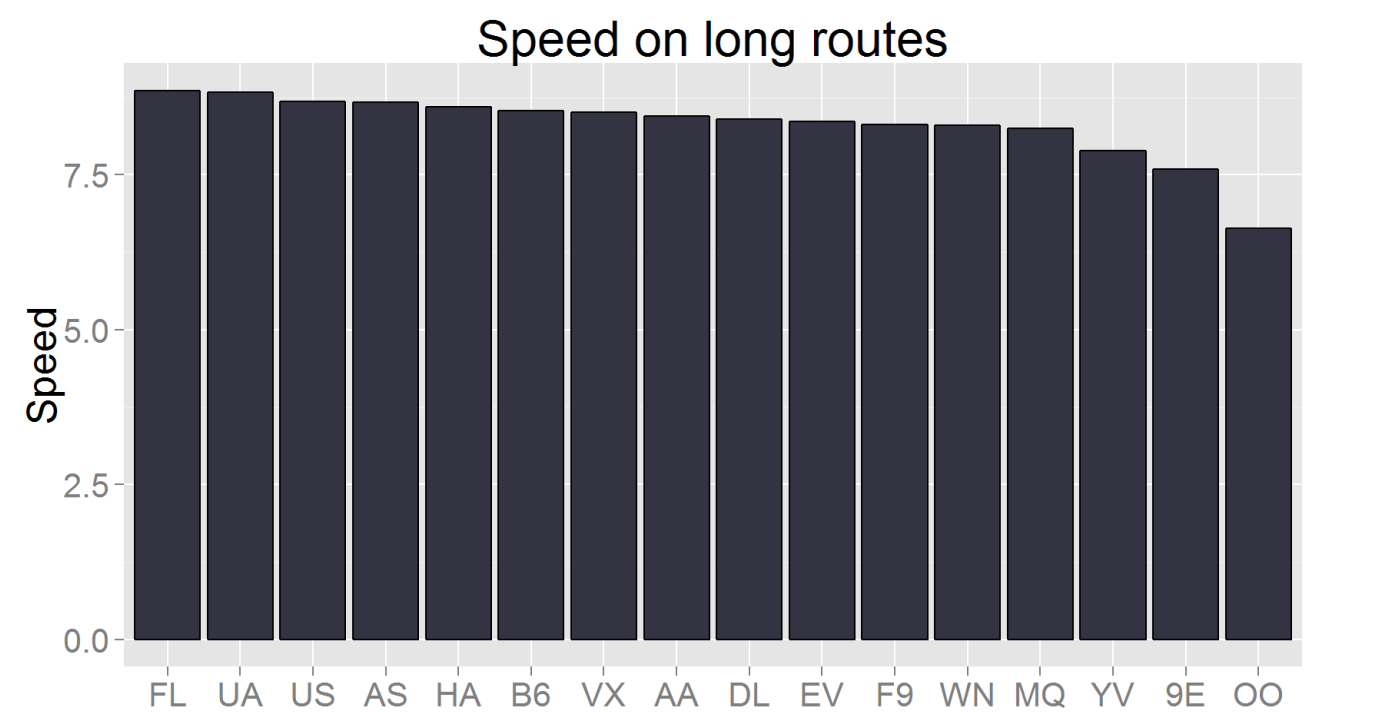


Here in column “Distance” we have the longest distance for each airline, in column “Time” the least time to overcome it.

##	ID	NAMe	Time	Distance	Speed
## 1	OO	SkyWest Airlines Inc.	152	1008	6.631579
## 2	9E	Endeavor Air Inc.	209	1587	7.593301
## 3	YV	Mesa Airlines Inc.	69	544	7.884058
## 4	MQ	Envoy Air	139	1147	8.251799
## 5	WN	Southwest Airlines Co.	257	2133	8.299611

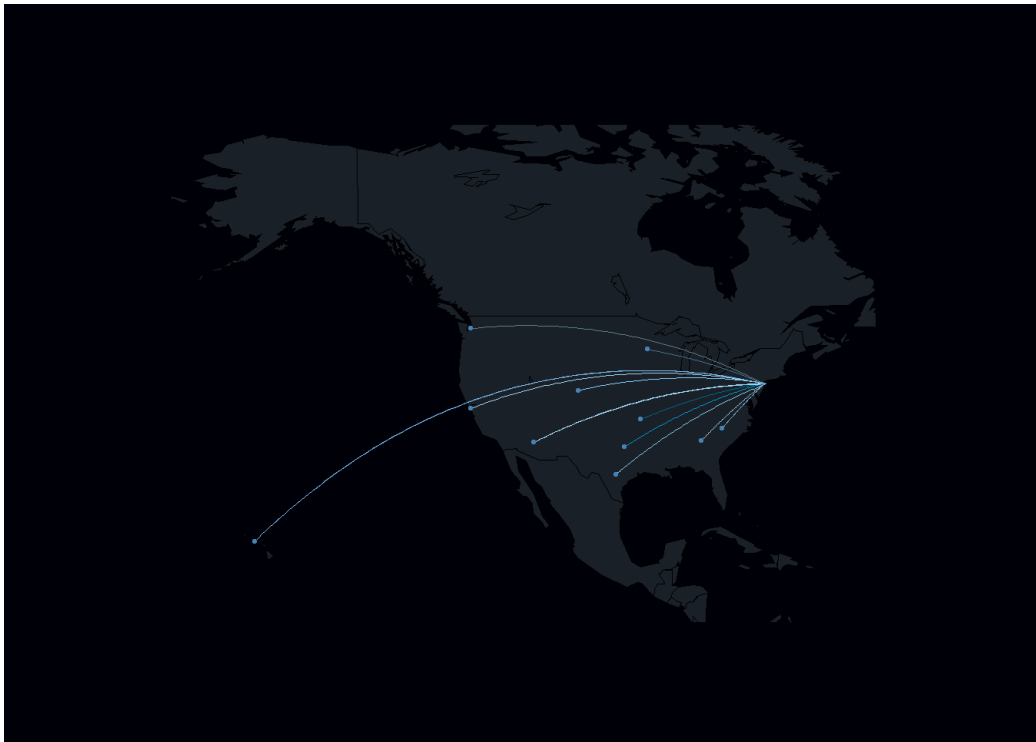
##	6	F9	Frontier Airlines Inc.	195	1620	8.307692
##	7	EV	ExpressJet Airlines Inc.	166	1389	8.367470
##	8	DL	Delta Air Lines Inc.	308	2586	8.396104
##	9	AA	American Airlines Inc.	306	2586	8.450980
##	10	VX	Virgin America	304	2586	8.506579
##	11	B6	JetBlue Airways	303	2586	8.534653
##	12	HA	Hawaiian Airlines Inc.	580	4983	8.591379
##	13	AS	Alaska Airlines Inc.	277	2402	8.671480
##	14	US	US Airways Inc.	248	2153	8.681452
##	15	UA	United Air Lines Inc.	562	4963	8.830961
##	16	FL	AirTran Airways Corporation	86	762	8.860465

Lets see it on the plot.



After analyzing our table and plot, clearly visible the best and worst airlines in this competition. Undoubted leaders are “United Air Lines” and “AirTran Airways Corporation”. Loser is: “SkyWest Airlines”. Display this:

Longest routes for each airline



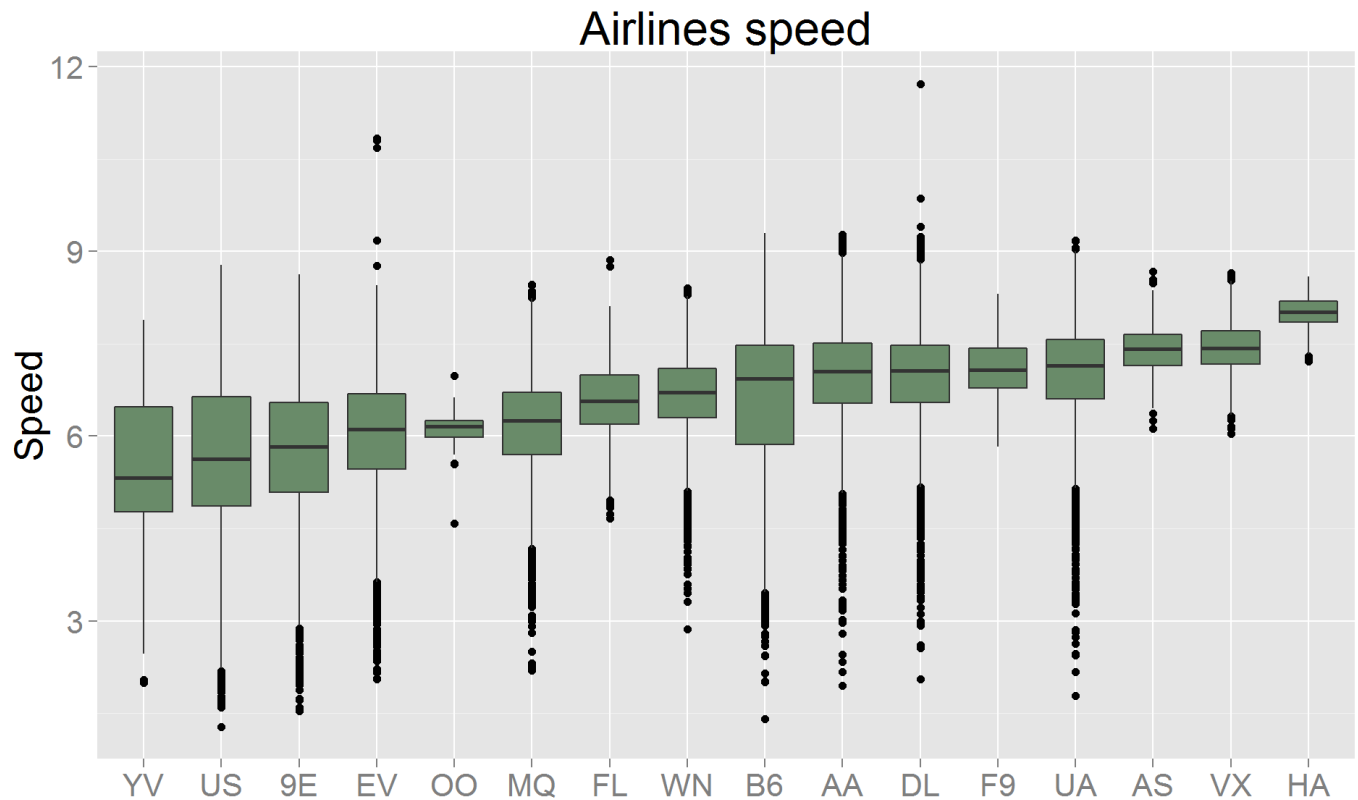
Now let’s see, how speed changes for each airline in general case.

##	Id	Name	speed
----	----	------	-------

```

## 1 YV          Mesa Airlines Inc. 5.726086
## 2 OO          SkyWest Airlines Inc. 6.100372
## 3 9E          Endeavor Air Inc. 6.106013
## 4 EV          ExpressJet Airlines Inc. 6.248766
## 5 MQ          Envoy Air 6.255462
## 6 US          US Airways Inc. 6.331736
## 7 FL AirTran Airways Corporation 6.572687
## 8 WN          Southwest Airlines Co. 6.744276
## 9 F9          Frontier Airlines Inc. 7.055776
## 10 B6         JetBlue Airways 7.075735
## 11 AA         American Airlines Inc. 7.113990
## 12 DL         Delta Air Lines Inc. 7.127570
## 13 UA         United Air Lines Inc. 7.230330
## 14 AS         Alaska Airlines Inc. 7.376747
## 15 VX         Virgin America 7.416662
## 16 HA         Hawaiian Airlines Inc. 7.997269

```



This plot and table shows completely different results then before. Here, the best, fastest and the most comfortable is Hawaian airlines. We can make a conclusion, that estimating airlines only on long routes isn't right.

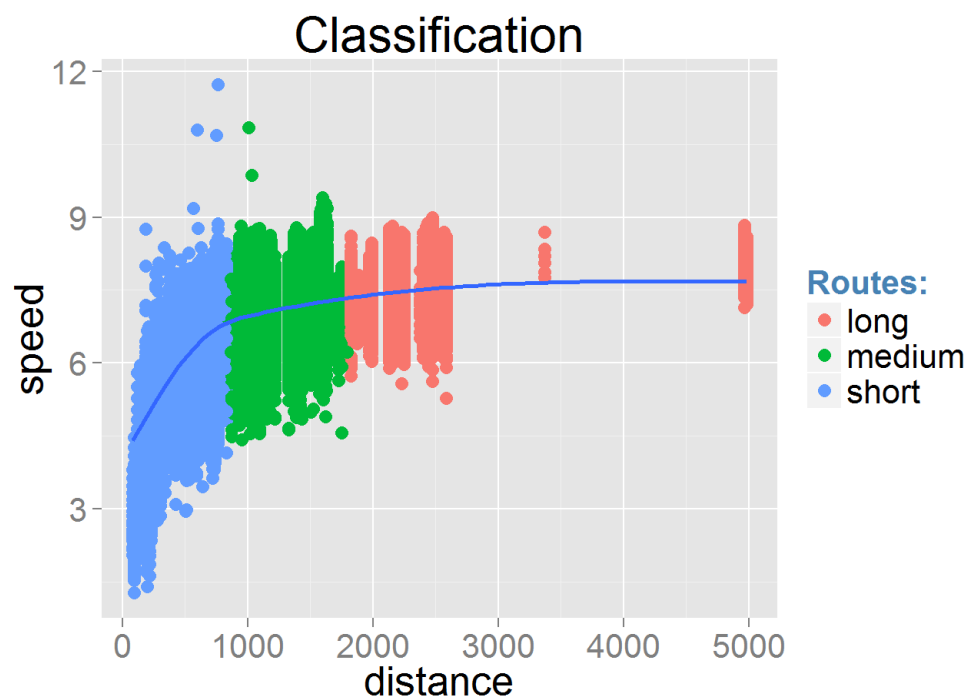
Dependence between speed and distance. Classification on groups. Clustering.

```

##
##
##      9E      AA      AS      B6      DL      EV      F9      FL      HA      MQ      OO
## long      0  6295    709   8417  10778      0      0      0    342      0      0
## medium 2688 18378      0  28073  21140  9157    681      0      0  2322      4
## short 14606  7274      0  17559  15740 41951      0   3175      0 22715     25
##
##      UA      US      VX      WN      YV
## long 19621  2240   5116    425      0
## medium 26132      0      0   4788      0
## short 12029 17591      0   6831    544

```

Table shows. which of the category of routes airline more prefer.



As we see, at the plot are visible three categories of routes. I tried to define, on which routes specialises each airline, long, medium or local.

Problem 2

Arriving delay trend

The problem is to figure out which airline makes steps for decreasing arriving delay. We will be comparing average delay at the beginning year with delaying at the end of the year.

First row it is the average arriving delay for each airline during first three months, second - during last three months.

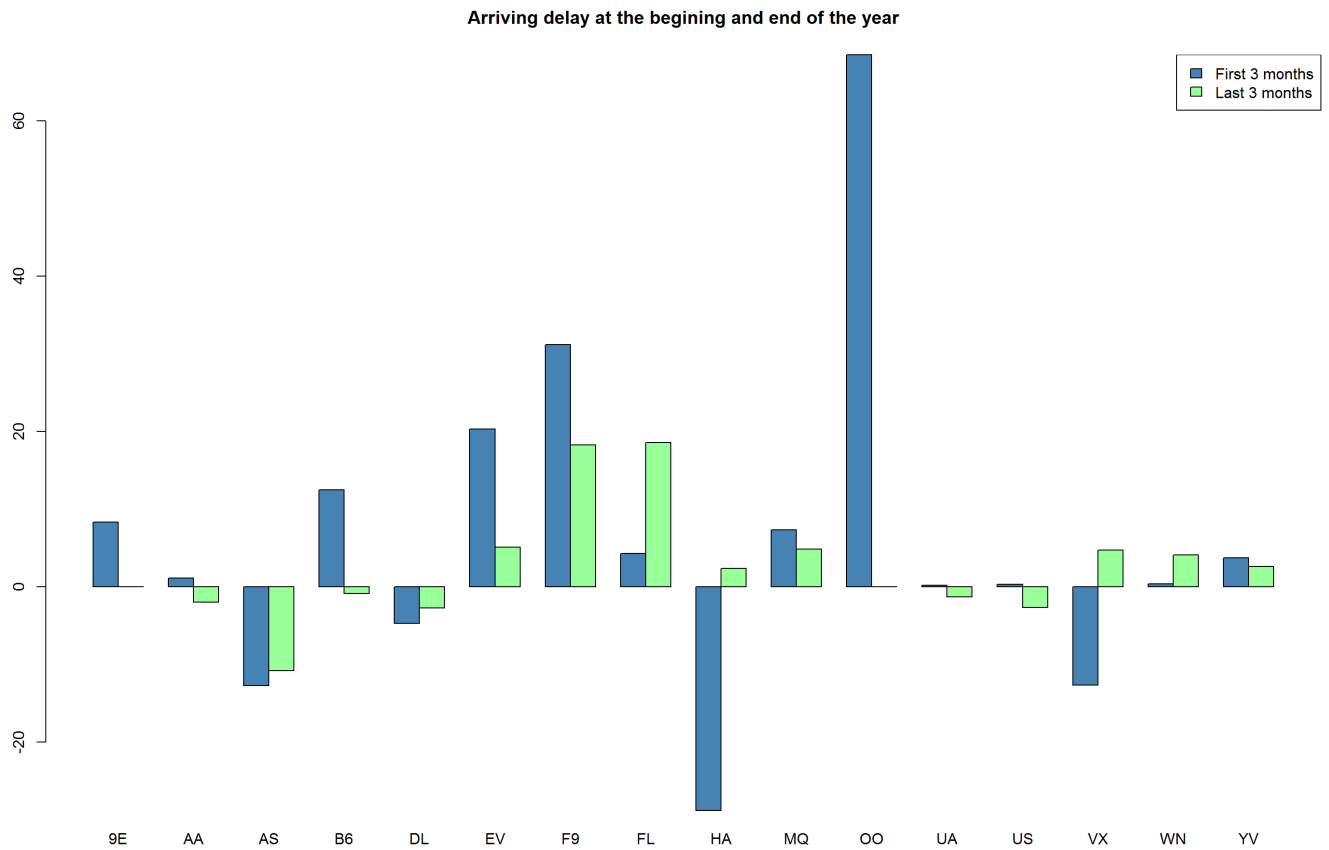
```
##      Month      9E      AA      AS      B6      DL      EV      F9      FL      HA      MQ      OO
## 2      1:3      8.28      1.09 -12.74 12.48 -4.73 20.27 31.15  4.25 -28.82  7.33 68.5
## 11 10:12 -0.03 -2.02 -10.81 -0.88 -2.76  5.09 18.22 18.54   2.32  4.85  0.0
##      UA      US      VX      WN      YV
## 2      0.19  0.27 -12.66  0.34  3.70
## 11 -1.30 -2.67   4.70  4.08  2.56
```

```
##      Month Airline  value
## 1      1:3      9E    8.28
## 2 10:12      9E   -0.03
## 3      1:3      AA    1.09
## 4 10:12      AA   -2.02
## 5      1:3      AS  -12.74
## 6 10:12      AS -10.81
```

Under each airlines ID we have the value that describe how changed delay during whole year.

```
##      9E      AA      AS      B6      DL      EV      F9      FL      HA      MQ
## Delay_change -8.31 -3.11  1.93 -13.36  1.97 -15.18 -12.93 14.29 31.14 -2.48
##      OO      UA      US      VX      WN      YV
## Delay_change -68.5 -1.49 -2.94 17.36  3.74 -1.14
```

Let's see it visually.



So, what we can say here. We can see that “SkyWest Airlines” shows us very good and positive result. Here this airline show us the biggest decreasing in arriving delay. Should also be noted “ExpressJet Airlines”.

Problem 3

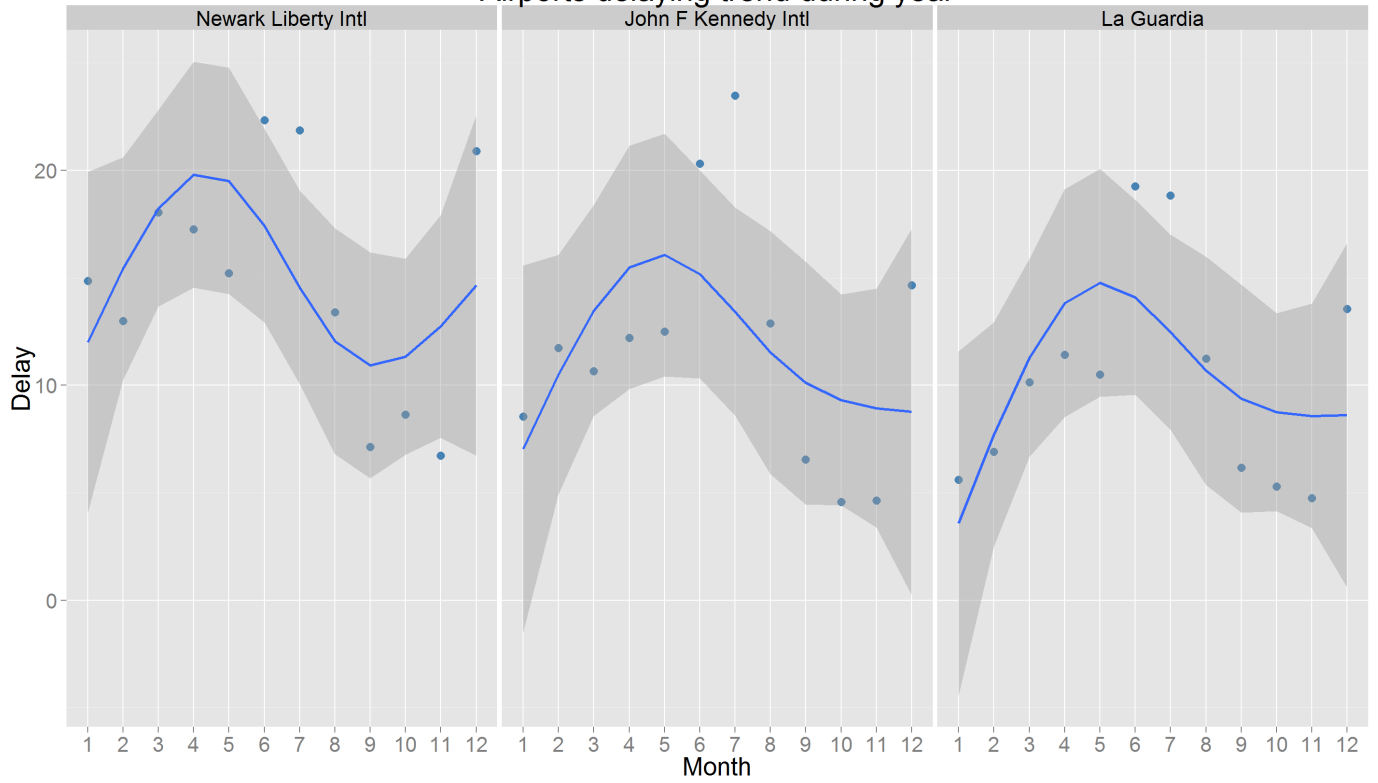
We have three airports. I'm going to make analysis for both arriving and departure delaying, to see trend of delaying, and how it changed during the year.

Origin departure delay

##	month	EWB	JFK	LGA
## 1	1	14.855657	8.557635	5.607147
## 2	2	12.995452	11.736605	6.921184
## 3	3	18.050624	10.670843	10.153993
## 4	4	17.255987	12.214801	11.430409
## 5	5	15.228768	12.495038	10.506955
## 6	6	22.338537	20.321934	19.254873
## 7	7	21.858977	23.479758	18.845541
## 8	8	13.408123	12.889260	11.238733
## 9	9	7.142170	6.553027	6.166027
## 10	10	8.636918	4.577946	5.287831
## 11	11	6.721649	4.643378	4.757767
## 12	12	20.911902	14.650454	13.565212

Let's look how it looks on plots.

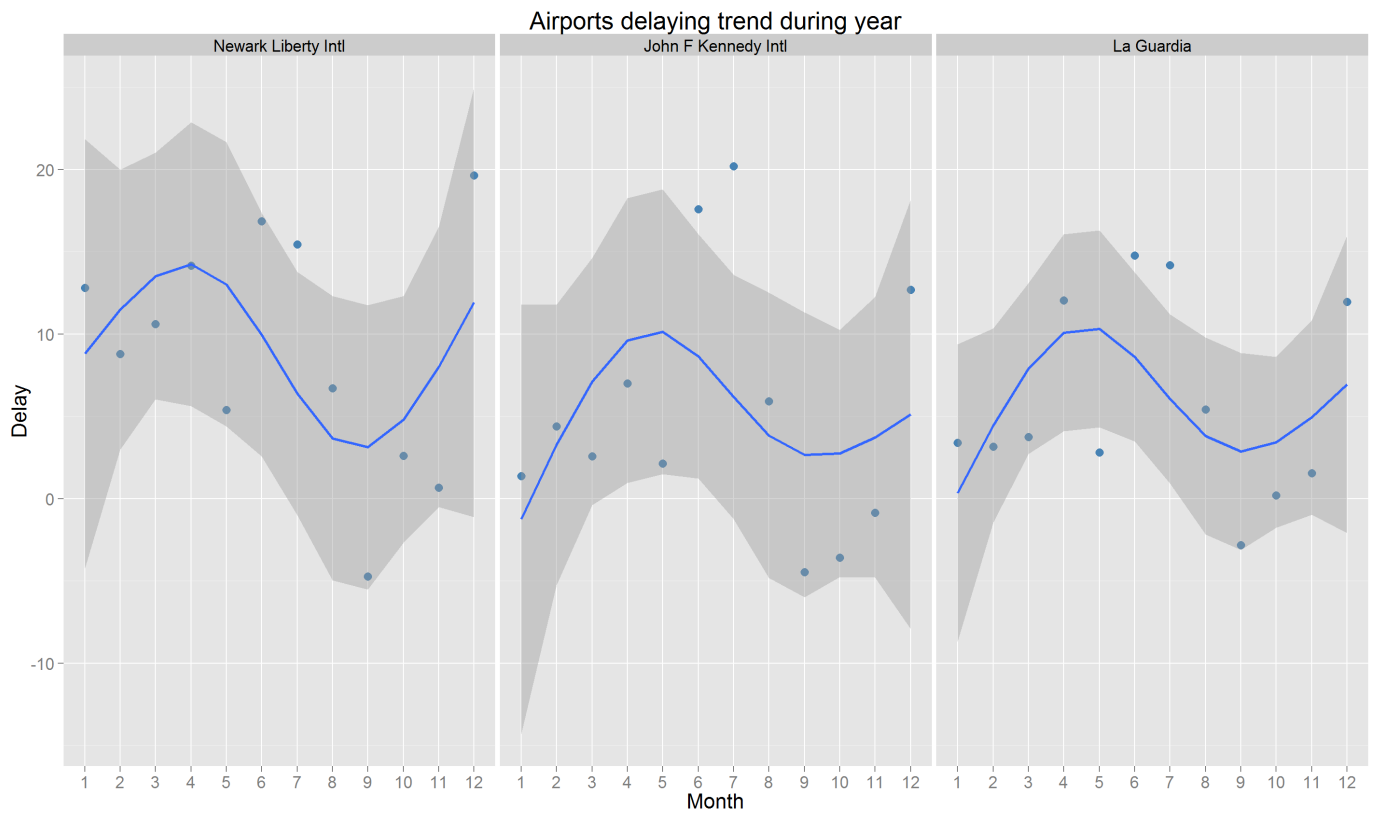
Airports delaying trend during year



Origin arriving delay

Now let's calculate all the same for arriving delay.

##	month	EWR	JFK	LGA
## 1	1	12.8165557	1.3683977	3.3824023
## 2	2	8.7751603	4.3910328	3.1473894
## 3	3	10.6007988	2.5808150	3.7384982
## 4	4	14.1433877	7.0115389	12.0385817
## 5	5	5.3819276	2.1229773	2.7963764
## 6	6	16.8635990	17.5969288	14.7692779
## 7	7	15.4602015	20.1902224	14.1815696
## 8	8	6.7123423	5.9108409	5.4078014
## 9	9	-4.7299722	-4.4630178	-2.8253950
## 10	10	2.6047372	-3.5859719	0.1864229
## 11	11	0.6724982	-0.8728745	1.5511865
## 12	12	19.6397450	12.6775748	11.9563716



Conclusion. We can see that in both cases in summer period of year we have a big increasing of delays. What causes this? Answer very simple. In this period of year in several times growing amount of tourist, and it's causes delays. What's interesting that in both cases present big decreasing of delays in autumn, it can be explained that this period of year is not so interesting for tourists to have trips and travelling, thats why airports isn't so loaded in autumn, and stuff have enough time to prepare passages for flying. So, it may be useful for summer time increase number of stuff in every airports and decrease it's for autumn, it must be profitable.