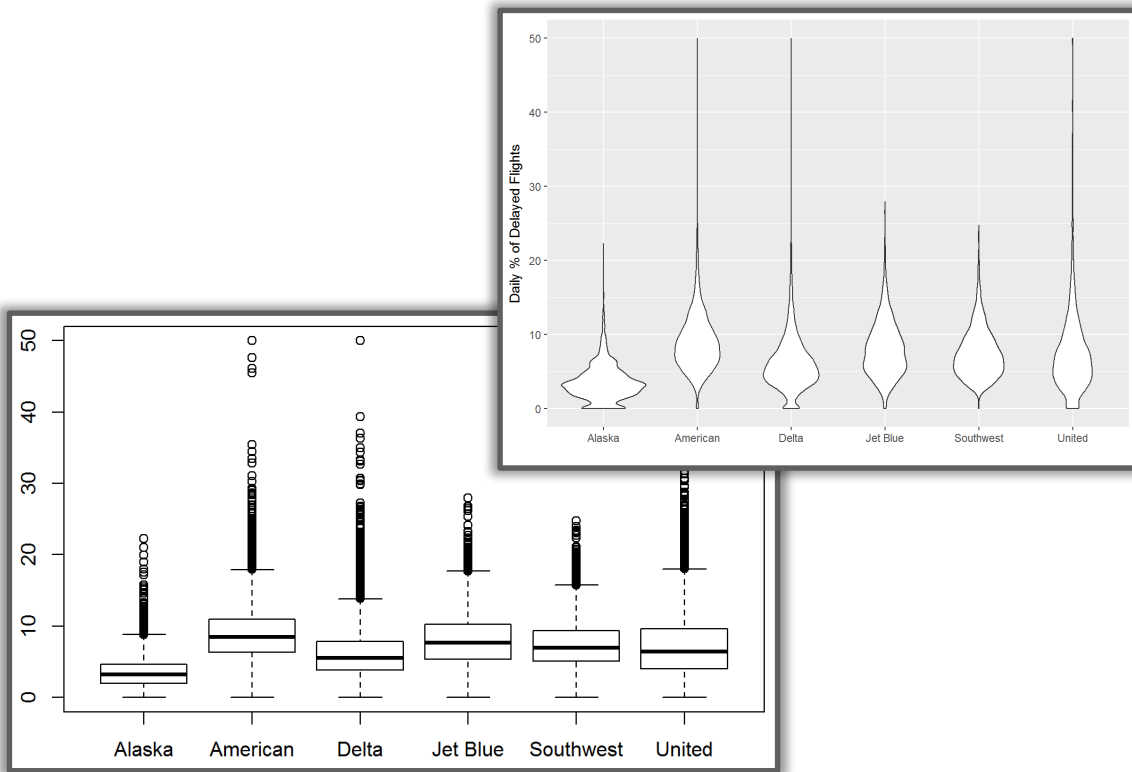


Bivariate EDA: Comparing Quantitative Variable against Categorical Variable



- Group-by Scatterplot Matrix
- Boxplots
- Violin plots of density of quantitative variable against a categorical variable

Prof. Nagiza F. Samatova

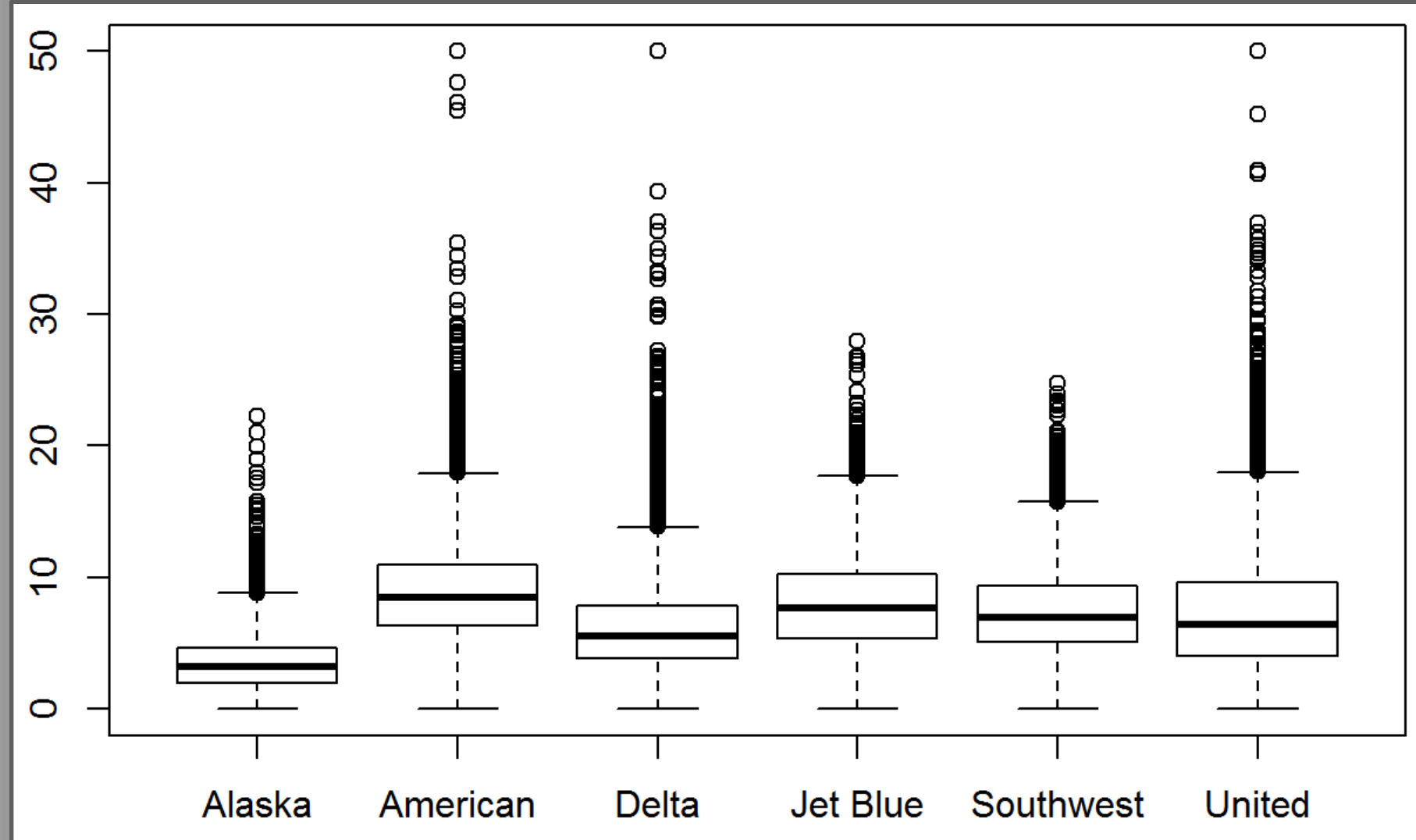
samatova@csc.ncsu.edu

Department of Computer Science
North Carolina State University

Bivariate Analysis: **Categorical** & **Quantitative** Vars

Measure	Description
Violin plot	similar to a boxplot but showing the density estimate
Scatterplot matrix	plot in which the x-axis is the value of one variable, and the y-axis is the value of the other variable; but grouped-by values of the categorical variable
Boxplot	plot to visualize distribution of data grouped by the values of the categorical variable

Boxplot: Quantitative against Categorical Data



Violin Plot: Density against Categorical Variable

```
1 import seaborn as sns
2 sns.violinplot(x="airline",
3               y="pct_carrier_delay",
4               data=airlines)
5 plt.show()
```

- Plots the density estimate
 - with the density on the y-axis
- Density is mirrored and flipped: **violin**
- Pros: Unlike boxplots, they can show **nuances in the distribution**
- Cons: Boxplot more clearly shows **outliers**

