* Import the package nycflights13. Merge the flights and planes data into a new data frame. Answer the following questions by the merged data.
* Regress arr\_delay on log(distance), seats, and origin. (1) Which airport is the base cateory? (2) explain the meaning of the coef. of log(distance), and (3) is the coef. of seats significantly different from -0.5 under the 5% significance level?
* Regress arr\_delay on log(distance), seats, and carrier. Which carrier is the worst carrier (namely, has the longest delay time)?
* Select only two carriers: AA and DL. Randomly sample 100 observations. Use set.seed(777) to fix the random seed.
* Creat the scatterplot of arr\_delay (y-axis) and dep\_delay (x-axis) based on the random sample from the previous step. (1) Color the points by the carrier. (2) Add a single regression line. (3) Label the destination. (4) The size of the points depends on the variable seat. (5) Apply the Wall Street Journal theme.