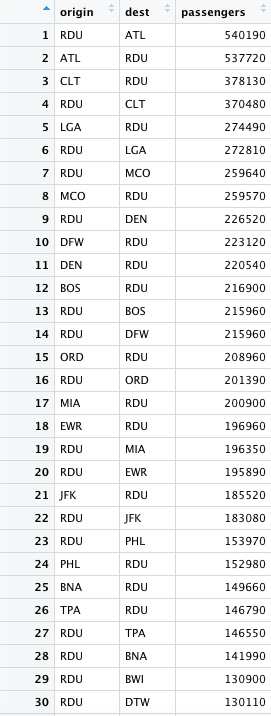
Sarah Followill

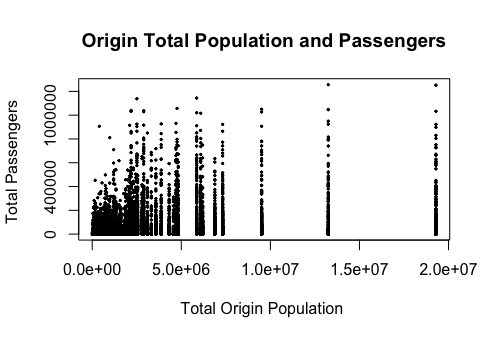
PLAN 372 – HW4 Report

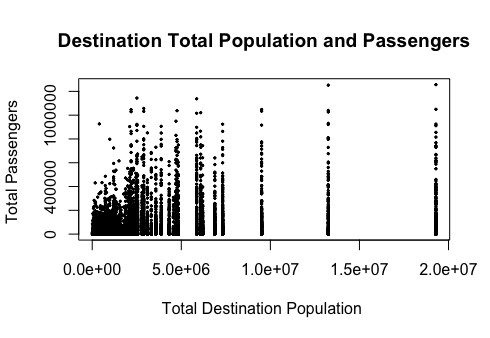
1.

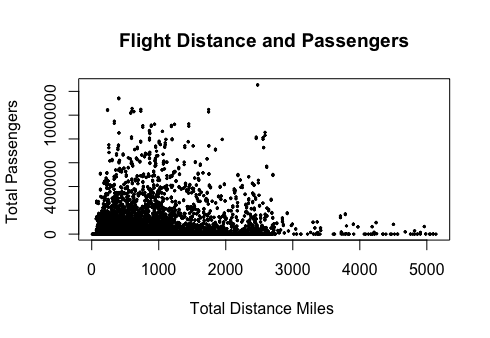


According to my summary table (first 30 rows shown above), the most popular nonstop destination from RDU is ATL, or Hartsfield Jackson Airport in Atlanta, Georgia. This flight pairing from RDU to ATL has a total of 540,190 passengers from October 2021 – September 2022.

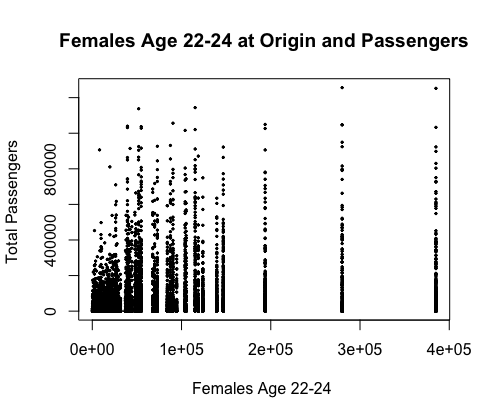
2.



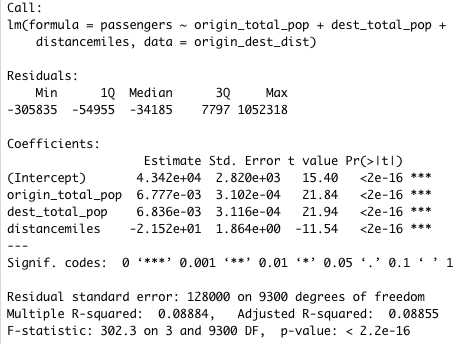




2. Extra Credit

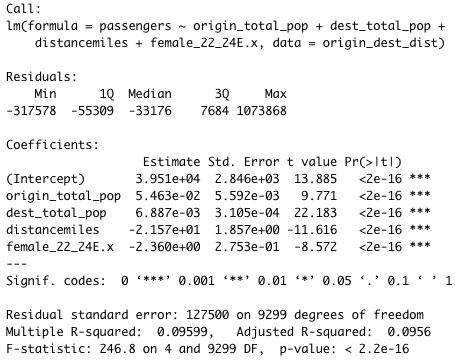


3.



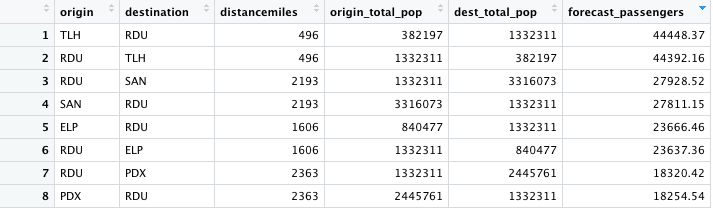
Above are the results of the multiple linear regression evaluating the effects of origin CBSA total population, destination CBSA total population, and total flight miles on the number of passengers on a given domestic flight. According to my results, a 1000-person increase in the origin city population will have a 6.78 passenger increase. A 1000-person increase in the destination city population will have a 6.84 passenger increase. A 1-mile increase in the total flight distance will decrease the passenger count by 2.15. These results make sense due to more people needing to fly out of a city’s airport if there are 1,000 more people in that city. The same logic holds for the destination city, as there will be more people to return home by plane in correspondence with an increase in destination city population. Finally, the longer the flight is, the fewer passengers will want to take that flight, so the distance coefficient makes sense as well. These results are statistically significant based on the low p-value of 2.2e-16, but the model is not a great fit due to the low R^2 value of 0.09, which is much closer to 0 (worst fit) than 1 (best fit).

3. Extra Credit: Including Females Age 22-24 Effect on Passengers



In adding the census variable of females ages 22-24 in the origin location and evaluating its effect on flight passengers, the coefficient appears to be a bit confusing. It is -2.360e+00, and anything to the 0th power is 1, so the coefficient is technically 1, meaning that there would be a 1:1 ratio between females ages 22-24 in the origin and additional passengers on a theoretical flight. The R^2 value for this multiple regression is actually slightly higher than the previous regression model, at 0.10, but this still means that the model is not a great fit. The p-value is the same as the previous model, indicating that these results are statistically significant.

4.



According to my passenger prediction model using my multiple regression model, Tallahassee, Florida (TLH) to Raleigh-Durham, NC (RDU) appears to be the flight that will be the most popular, seconded by the reciprocal flight from RDU to TLH. However, based on the extremely low R^2 in my multiple regression model, meaning that the model is not the best fit for the data, this prediction model might not be the most accurate. None of the above results appear to be nonsensical, but they might not be entirely accurate estimations based on the low R^2 value of my multiple regression.