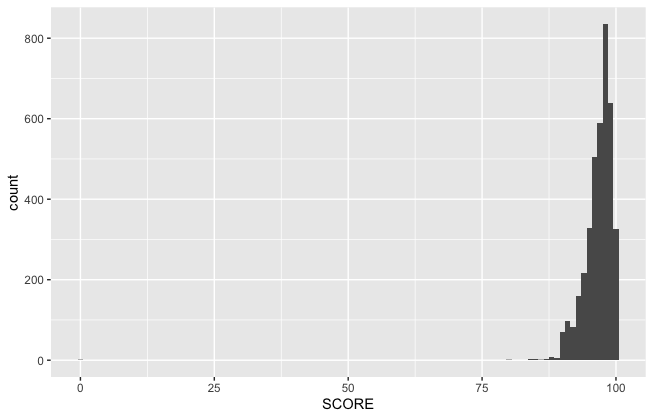
Restaurant dataset

Sarah Followill

LINK TO HW2 REPOSITORY: <https://github.com/sarahfollowill/plan372_hw2>   
   
The restaurant dataset contains records of the most recent health inspection for food-service establishments in Wake County. County health officials are curious to gain a better understanding of the overall picture of food safety in the county, in order to better target enforcement efforts, and have asked you use this dataset to answer the following questions:

1. Visualize the overall distribution of inspection scores using a histogram. [1 point]

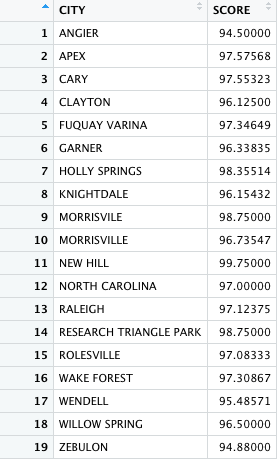


2. Some restaurants have been in business much longer than others. Is there any trend in terms of how highly older vs. newer restaurants score on their inspections? [0.5 points]



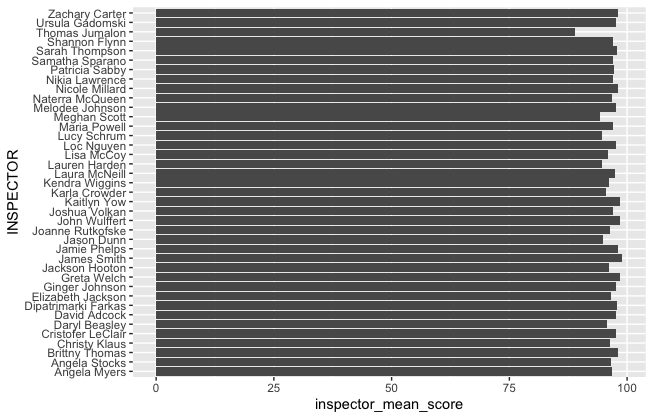
No, there is no real trend between the date that the restaurant opened and its scoring, as seen by the above chart. Older restaurants towards the left are scoring within the same range as the newer restaurants towards the right. There is one inspection listed in the data where no restaurant opening date is listed, which was removed. It should not affect the spread of results considering it is only one inspection amongst thousands that do have dates.

3. Wake County is the most populous county in North Carolina, and there are many cities in it. Do the inspection scores vary by city? Note that the city column contains some differently spelled city names; make sure to clean those up so that there is only one estimated value per city. The recode function that we used for creating a weekend/weekday variable in the SFpark exercise will be useful here, and you may also be interested in the str\_to\_upper function. [1 point]



Inspection scores do vary moderately by city, but every city in Wake County’s average restaurant score is between 94 – 100. Therefore, there is not too wide of a variation in scores amongst the cities. There are 295 inspections conducted that do not have a city named in the data. This number is somewhat significant and therefore could affect the various mean scores for each city as well as the range of mean scores. There is also one inspection where the city listed just as “North Carolina”, which I kept in my final data spread as it represents a mobile food unit that I assume goes to multiple cities in the area.

4. Wake County employs a whole team of inspectors. It is possible that some inspectors may be more thorough than others. Do inspection scores vary by inspector? [0.5 points]



Inspection scores do vary moderately by inspector, as seen in the above chart of average scores by inspector. Most of the inspector mean scores are in the same general range, between 95 – 99, but there are some outliers, like Thomas Jumalon, who has an average inspection score of 89.0, likely indicating that he is conducts extremely thorough inspections. On the other end of the spectrum, James Smith has an average inspection score of 99.0, likely indicating that this inspector is less thorough during his inspections.

5. It is possible that some extreme results from the previous questions are due to small sample sizes in a particular city, for a particular inspector, or in a particular time period. Look at the sample sizes in each of your groups. Do you think this is an explanation for the results you came to above? [0.5 point]

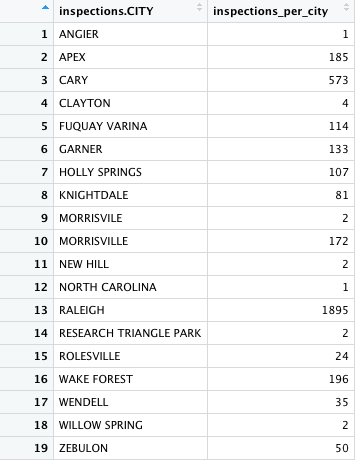


Fig. 5.1

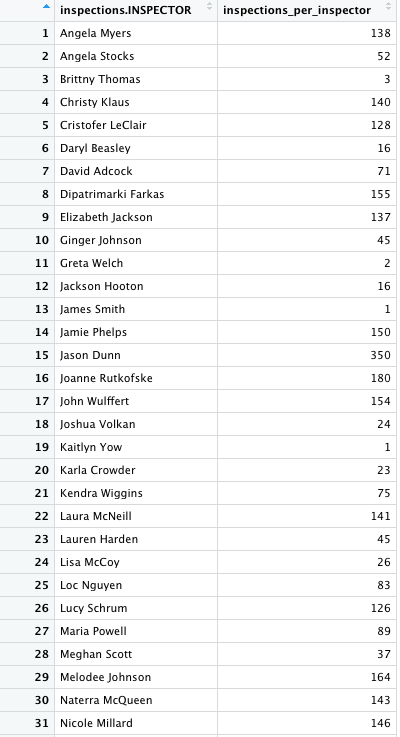


Fig. 5.2

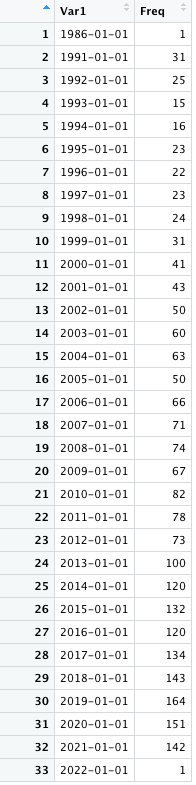
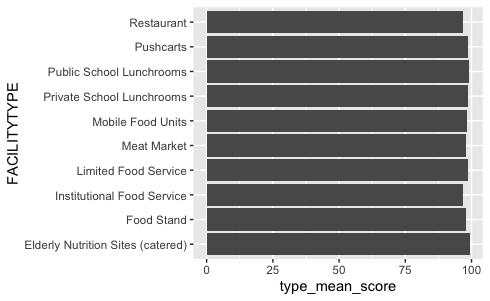


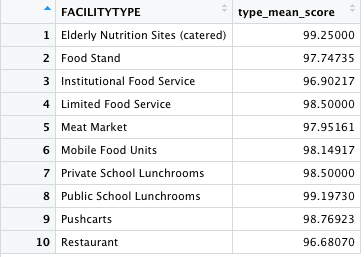
Fig. 5.3 *Count of inspections per year*

5. (cont.)

In the above questions, it is possible that small sample sizes could lead to disparities in results. For example, as seen in Fig. 5.1, some cities have had far fewer inspections conducted than others, as the city Angier only has one inspection while Raleigh has 1,895 inspections listed. For the cities, it is worth noting that some results might be skewed by the fact that there are 295 inspections listed that do not have cities associated with them in the data. Additionally, some inspectors have conducted many more inspections than others. For example, as seen in Fig. 5.2, Jason Dunn has completed 350 inspections that are listed, while James Smith only completed 1 that is listed. Finally, there are only 188 entries listed from restaurants opened before the year 2000, while there are 2,025 entries listed from restaurants opened from 2000 - 2022. There was one restaurant inspected that had no restaurant open date associated with it, but it should not affect the data much. The rest of these disparities could affect the results as the sample sizes have varied greatly in each of the categories evaluated.

6. The data file contains records for many types of food-service facility (e.g. restaurants, food trucks, etc.). Are the scores for restaurants higher than other types of facility? [0.5 point]

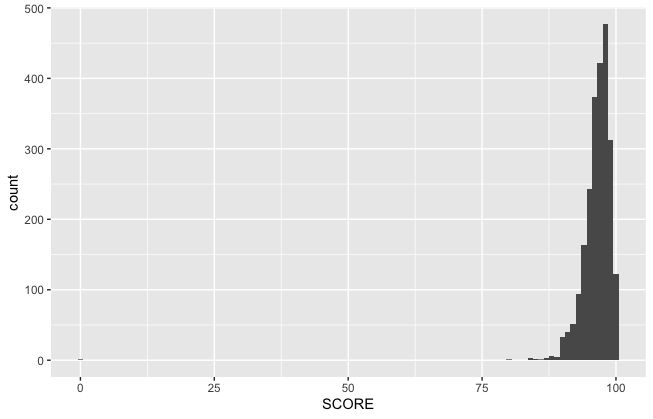




According to the above graph and table, average scores for restaurants are not higher than all other facility types. In fact, it has the lowest mean score of all facility types. There were 295 inspections listed in the original data that did not have a facility type associated in the listing. This disparity is somewhat significant and could affect mean scores if they were to be categorized into their respective facility types.

7. Since restaurants are where the general public is most likely to interact with the food-service system, Wake County Public Health is particularly interested in sanitation in restaurants. Repeat the analyses above (1-5) for restaurants specifically. [2 points]

7.1. Visualize the overall distribution of inspection scores for restaurants only using a histogram.



7.2. Some restaurants have been in business much longer than others. Is there any trend in terms of how highly older vs. newer restaurants score on their inspections?



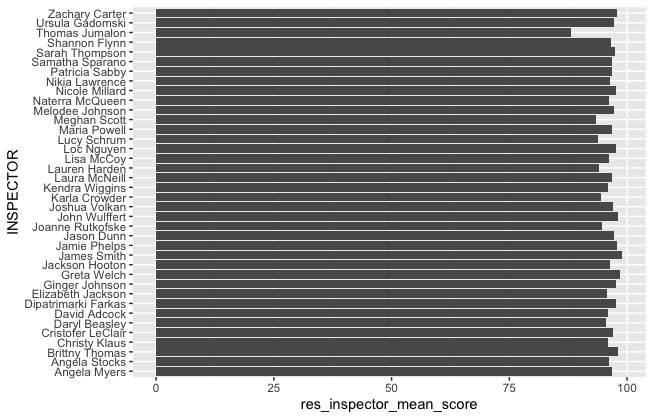
No, there is no real trend between the date that the restaurant opened and its scoring, as seen by the above chart. Older restaurants towards the left are scoring within the same range as the newer restaurants towards the right.

7.3. Wake County is the most populous county in North Carolina, and there are many cities in it. Do the inspection scores for restaurants only vary by city?



Inspection scores for restaurants do vary moderately by city, but every city in Wake County’s average restaurant score is between 93 – 100. Therefore, there is not too wide of a variation in scores amongst the cities.

7.4. Do inspection scores of only restaurants vary by inspector?



Inspection scores do vary moderately by inspector, as seen in the above chart of average scores by inspector. Most of the inspector mean scores are in the same general range, between 95 – 99, but there are some outliers, like Thomas Jumalon, who has an average restaurant only inspection score of 88.0, likely indicating that he is conducts extremely thorough inspections. On the other end of the spectrum, James Smith has an average restaurant only inspection score of 99.0, likely indicating that this inspector is less thorough, during his inspections.

7.5. It is possible that some extreme results from the previous questions are due to small sample sizes in a particular city, for a particular inspector, or in a particular time period. Look at the sample sizes in each of your groups. Do you think this is an explanation for the results you came to above for restaurants only?

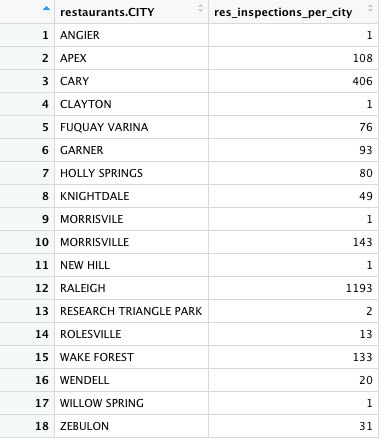


Fig. 7.5.1

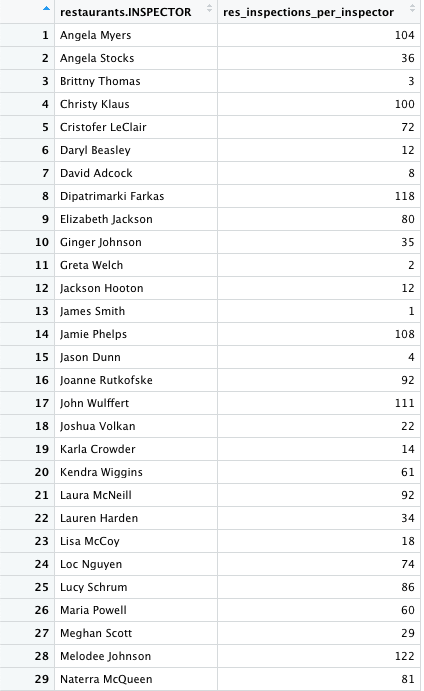




Fig. 7.5.2

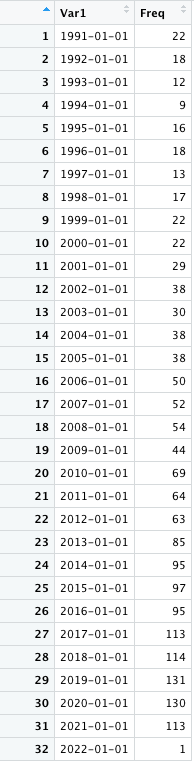


Fig. 7.5.3 *– Count of Restaurant Inspections per Year*

7. (cont.)

In the above questions, it is possible that small sample sizes could lead to disparities in results. For example, as seen in Fig. 7.5.1, some cities have had far fewer inspections conducted than others, as the city Angier only has one inspection while Raleigh has 1,193 inspections listed. Additionally, some inspectors have conducted many more inspections than others. For example, as seen in Fig. 5.2, Melodee Johnson has completed 122 inspections that are listed, while James Smith only completed 1 that is listed. Finally, there are only 147 entries listed from restaurants opened before the year 2000, while there are 1,565 entries listed from restaurants opened from 2000 - 2022. All of these disparities could affect the results as the sample sizes have varied greatly in each of the categories evaluated.