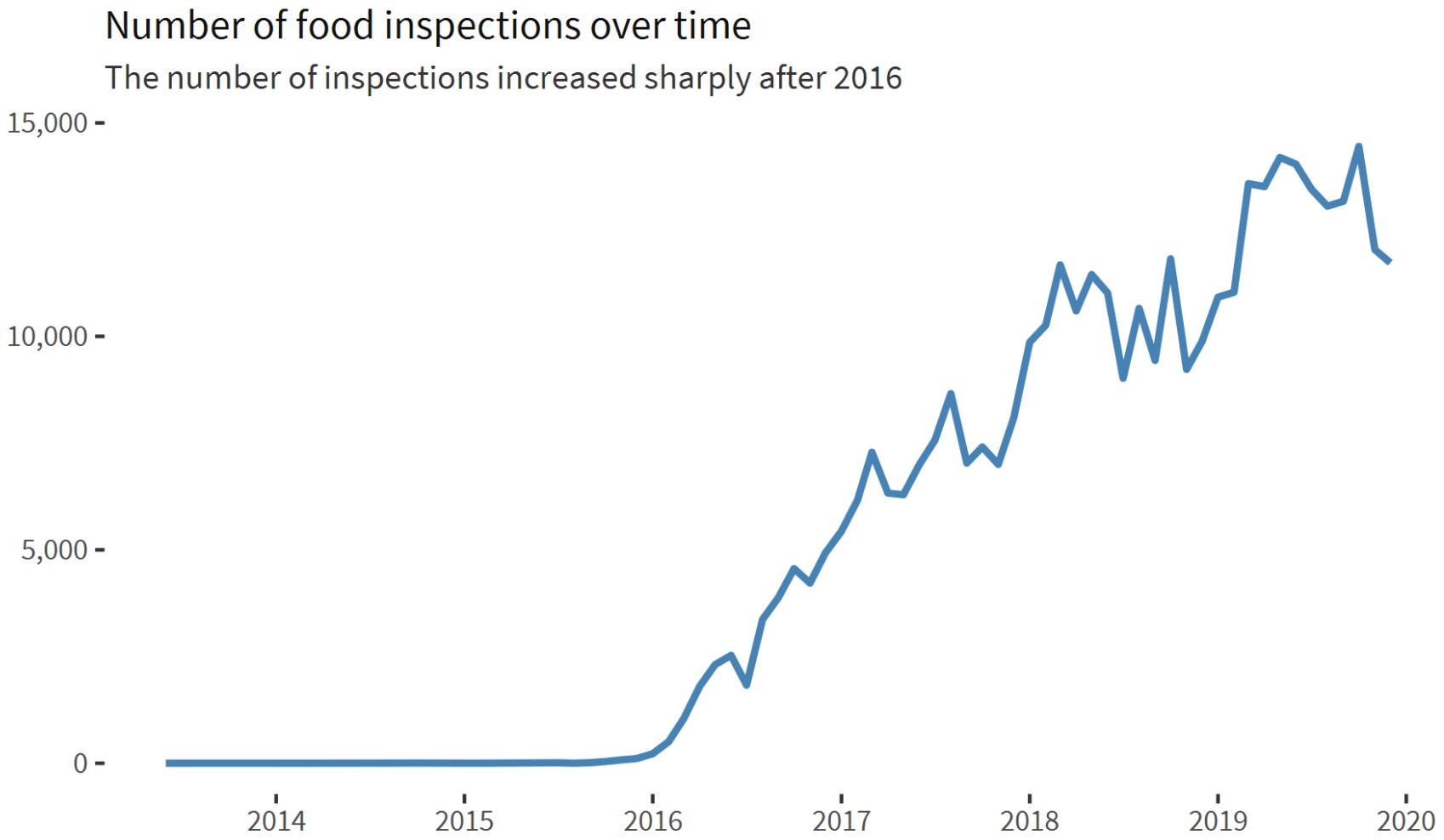
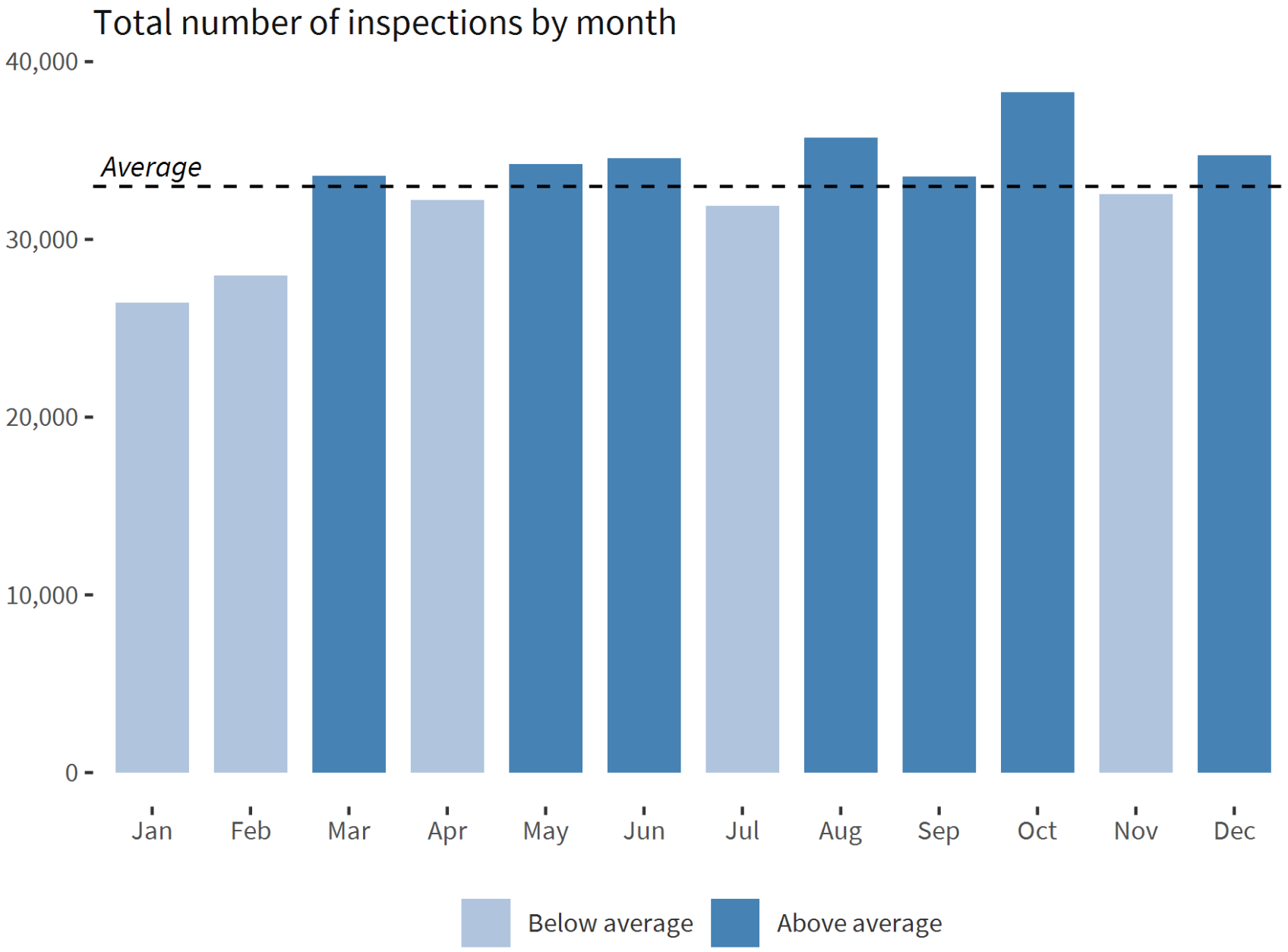
Question 1

How does the number of inspections change over time (use month as the level of temporal granularity)? Does the number of inspections increase or decrease over time? Are there any peak times? Is there any seasonal effect (like inspections being more common during certain seasons or months)?

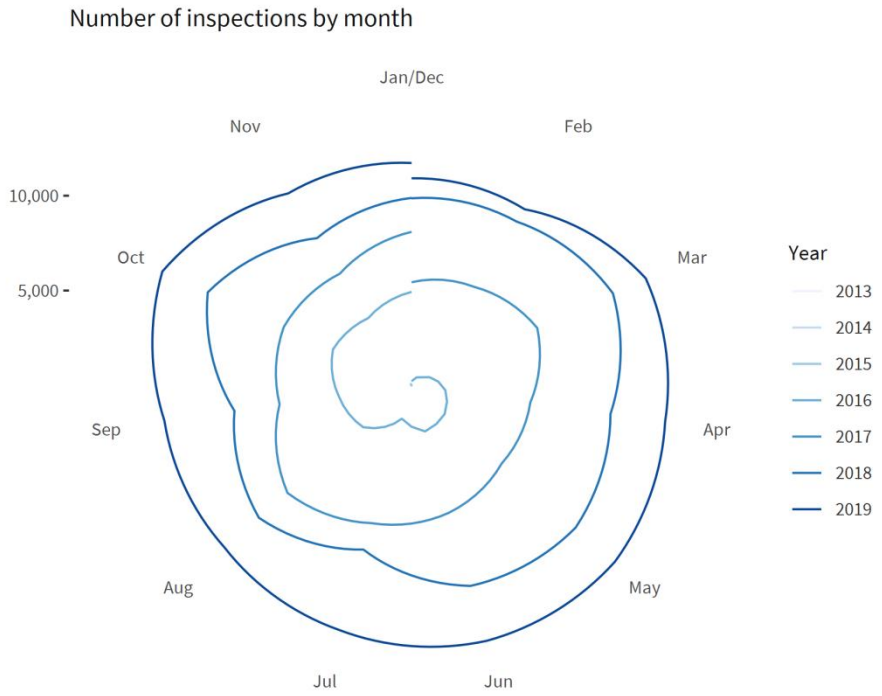

Number of food inspections over time
The number of inspections increased sharply after 2016
15,000
10,000
5,000
0
2014
2015
2016
2017
2018
2019
2020


Total number of inspections by month
40,000
30,000
20,000
10,000
0
Average
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Below average
Above average

Comments

The line chart shows the number of inspections by month, in a continuous scale. A line chart is very well suited in this case, since it allows us to perceive patterns in time.

To answer the question about seasonal effects, I first transformed the line chart to polar coordinates, since it would probably show any evident recurring pattern along the years:

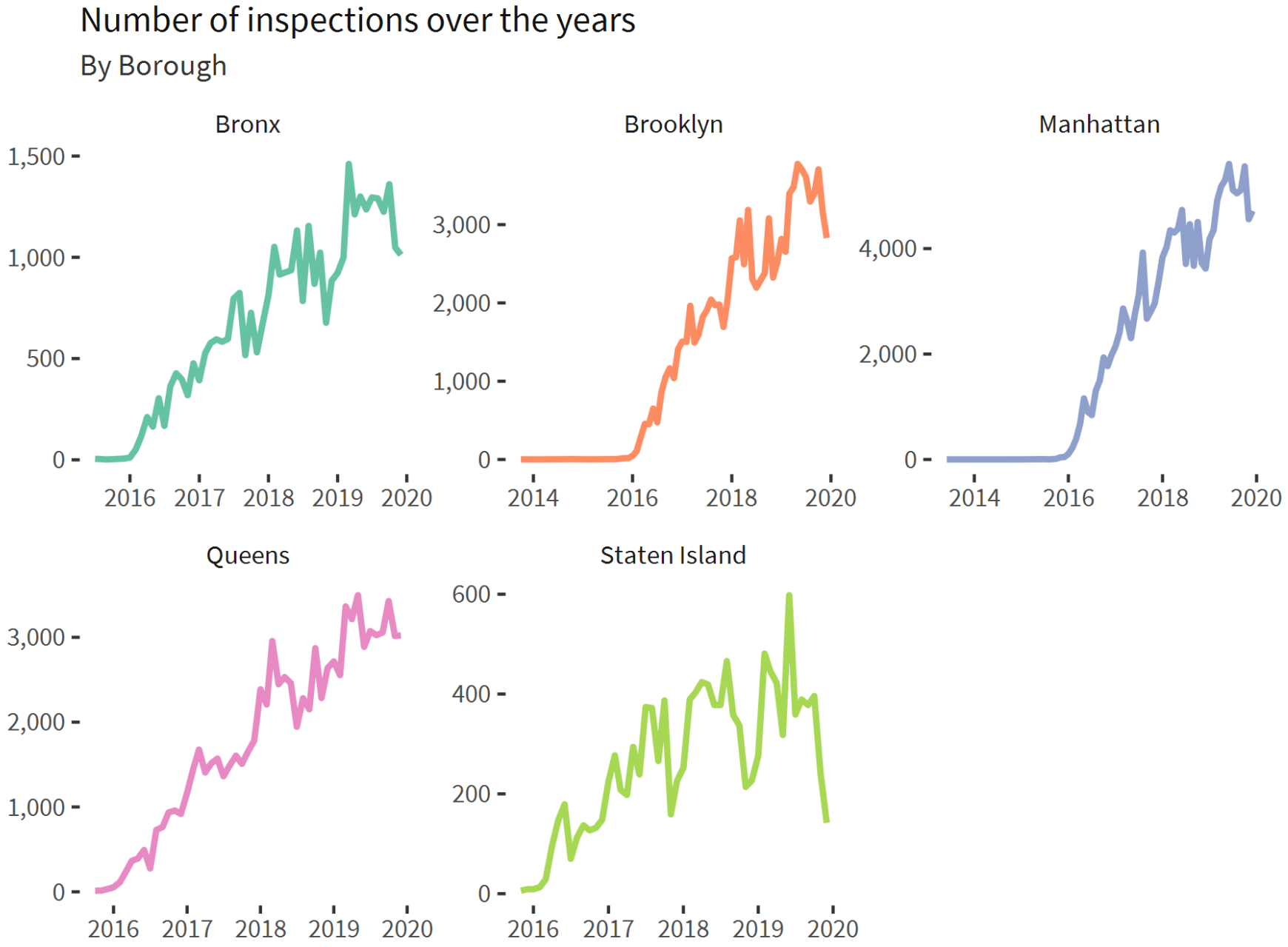


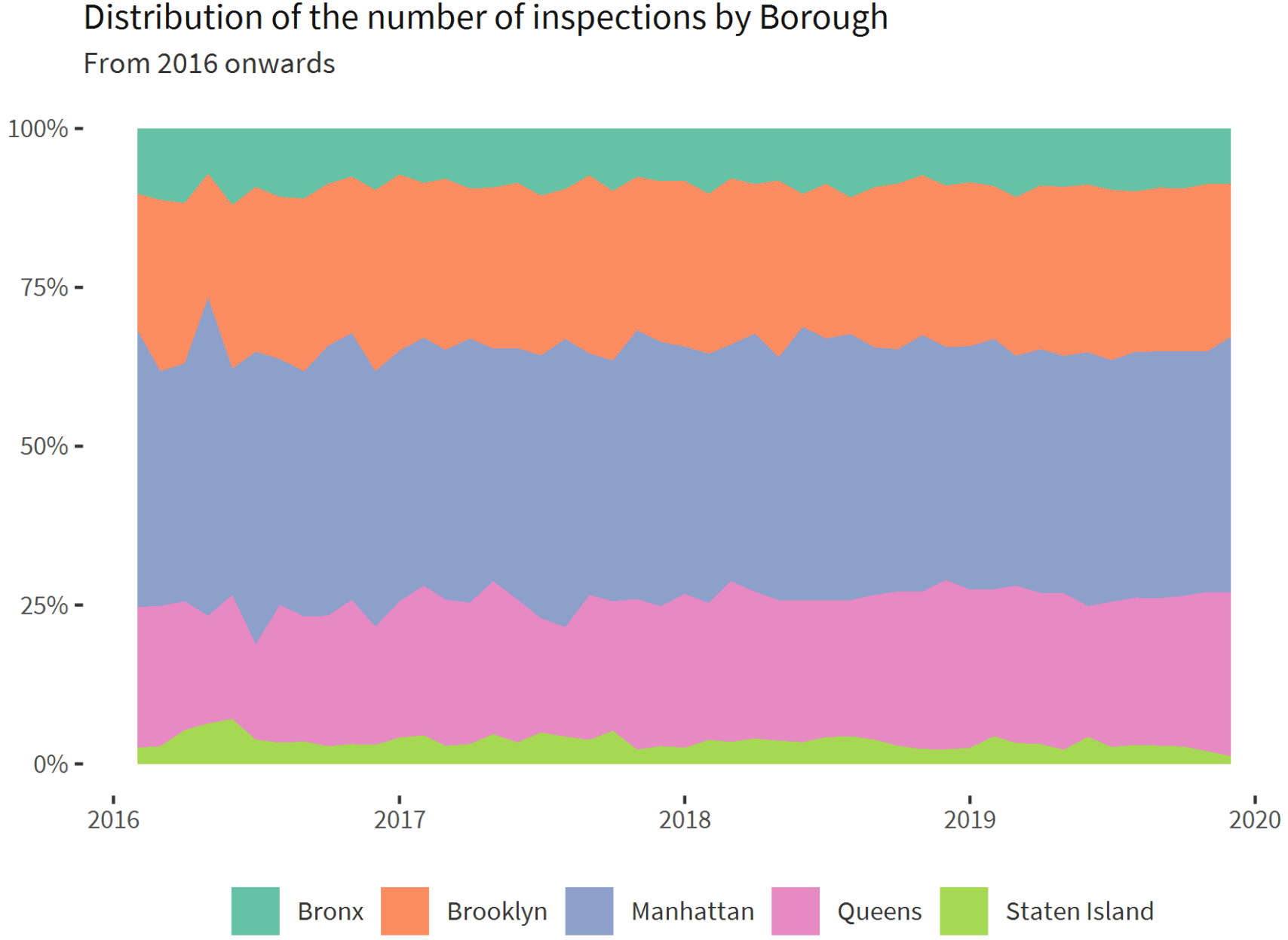
Comments

But there was no distinctive pattern that would “jump to the eye”, so instead I chose to design a more conventional bar chart, showing the total number of inspections by month, regardless of the year. Additionally, I included a line indicating the monthly average, highlighting the months above this average.

Question 2

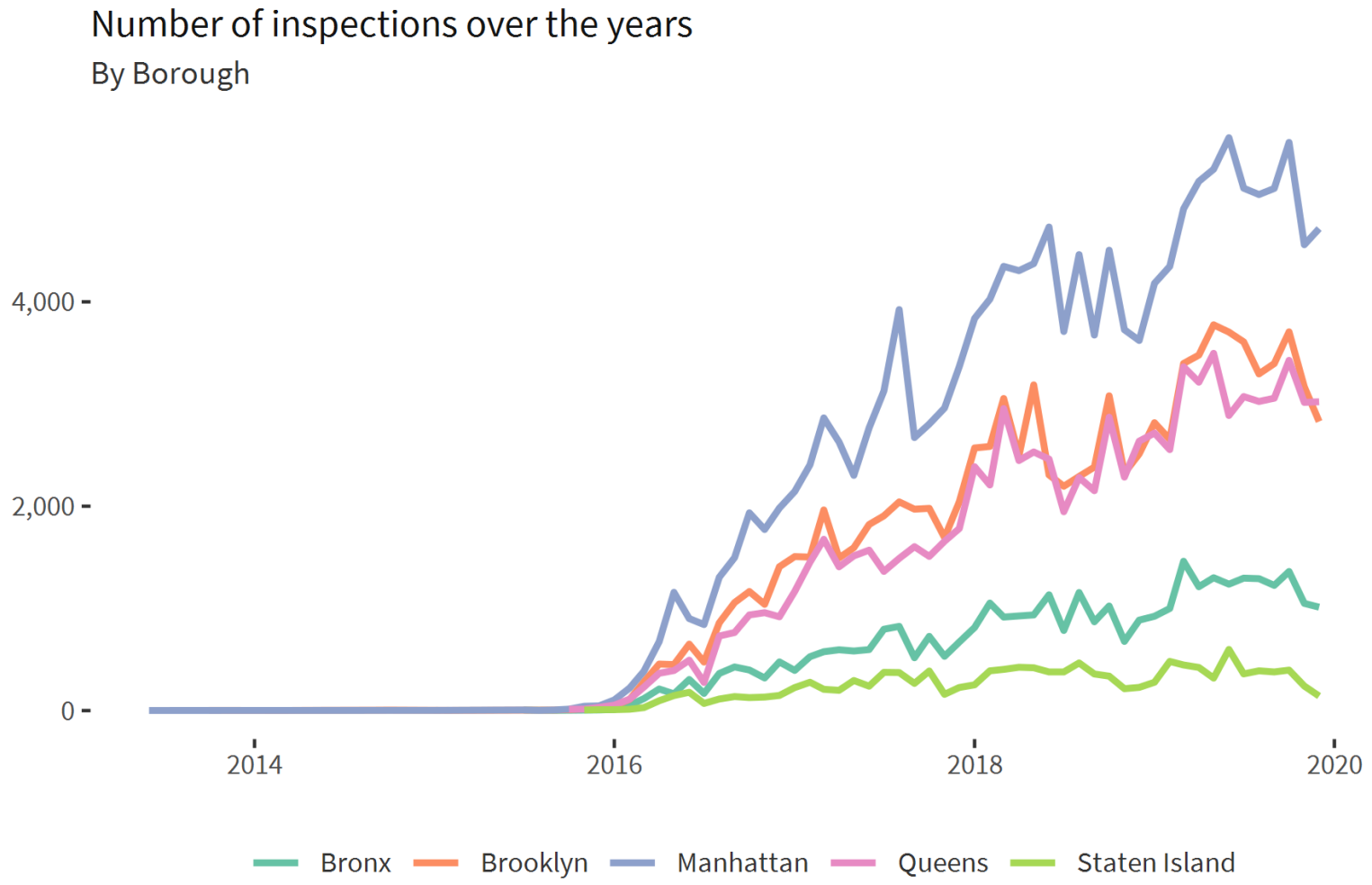
Is there any difference in how the number of inspections changes over time in the 5 different boroughs of New York City?


Number of inspections over the years
By Borough
Bronx
1,500
1,000
500
0
2016
2017
2018
2019
2020
Brooklyn
3,000
2,000
1,000
0
2014
2016
2018
2020
Manhattan
4,000
2,000
0
2014
2016
2018
2020
Queens
3,000
2,000
1,000
0
2016
2017
2018
2019
2020
Staten Island
600
400
200
0
2016
2017
2018
2019
2020


Distribution of the number of inspections by Borough
From 2016 onwards
100%
75%
50%
25%
0%
2016
2017
2018
2019
2020
Bronx
Brooklyn
Manhattan
Queens
Staten Island

Comments

Here, to show the patterns for each borough, the immediate idea would be to plot a line chart with multiple lines representing the boroughs :



Comments

But I actually in the end chose to use a faceted version of the chart, with free scales, to show that the pattern is very similar for every borough.

To demonstrate this further, I also plotted the distribution of the number of inspections by borough, to show that the percentage of inspections at each borough is more or less stable over the years. I disregarded the years prior to 2016, because the number of inspections in general clearly took off on this year (see chart on the previous slide).

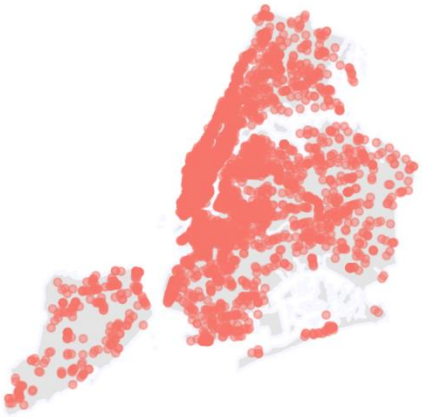
Question 3

How are cuisines types distributed across the New York area? Are there geographical areas where certain cuisines tend to concentrate (that is are there any areas where certain cuisines are more prevalent than others)? NOTE: focus only on the top 5 most frequent “Cuisine Description” categories.

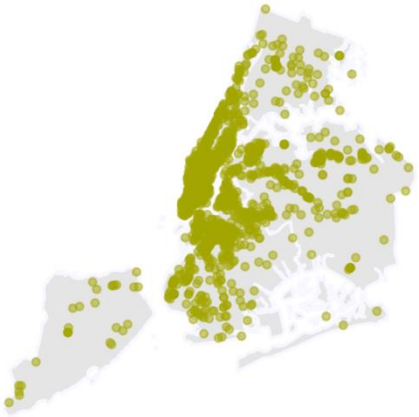
Inspections by cuisine in NYC area

Only top 5 cuisines shown

American

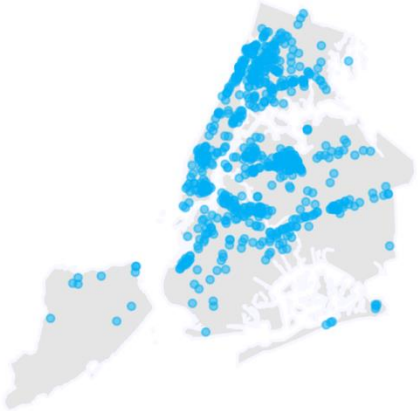


Café/Coffee/Tea

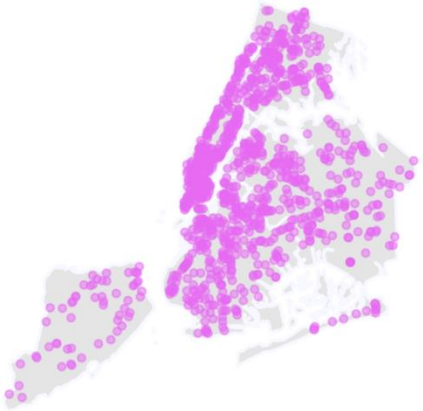


Chinese

Latin (Cuban, Dominican, Puerto Rican, South & Central American)



Pizza



Comments

Again I chose to use a faceted version of the chart, to avoid cluttering.

Question 4

How does the average score compare across different cuisine types? Are there cuisines that tend to have consistently lower/higher average scores compared to the others? NOTE: focus only on the top 5 most frequent “Cuisine Description” categories.


Distribution of the Inspections's scores
Among the top 5 cuisines
Average
Pizza
19.6
Latin (Cuban, Dominican, Puerto Rican, South & Central American)
22.3
Chinese
22.2
Café/Coffee/Tea
17.4
American
19.1
0
50
100
150
Score

Comments

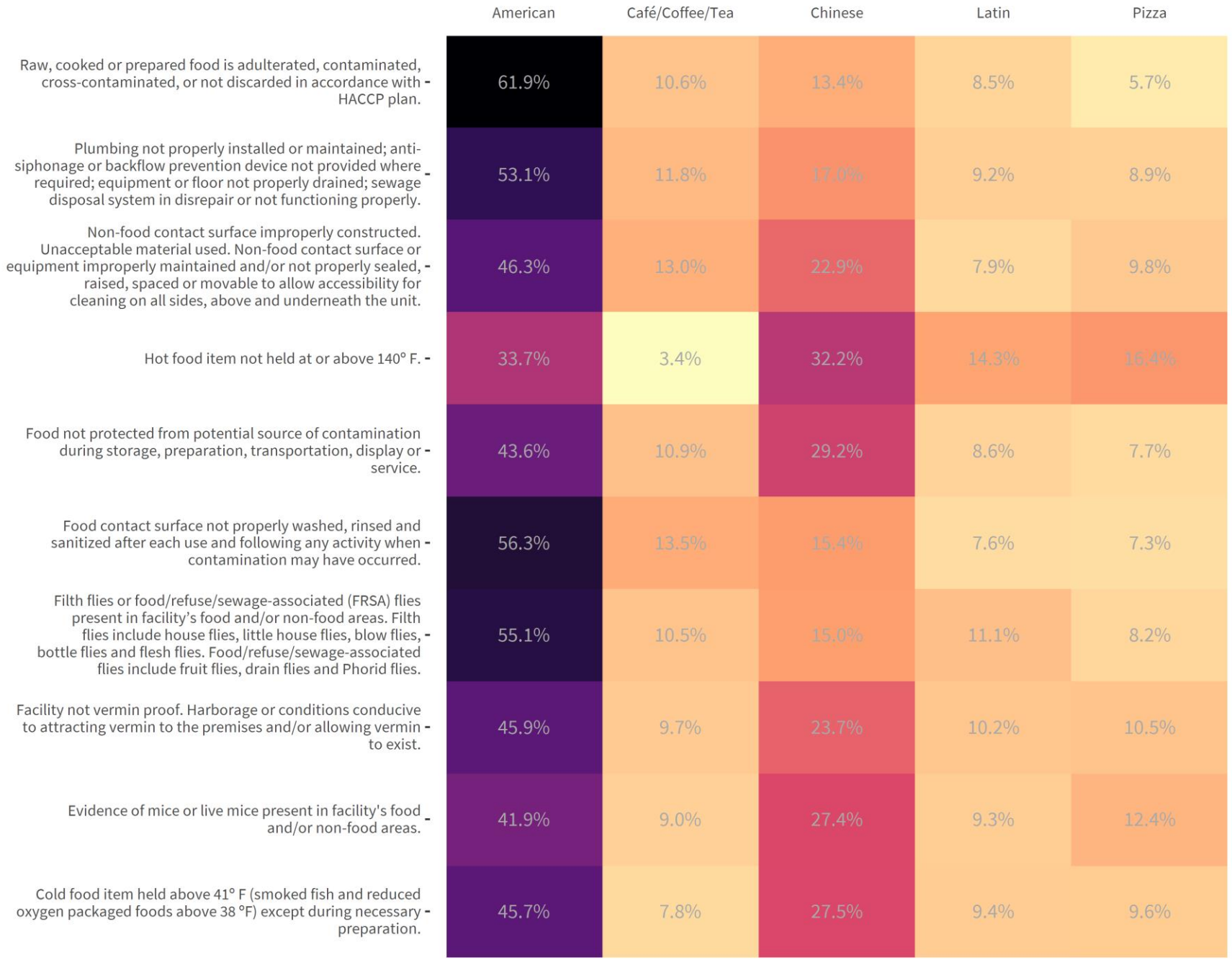
Instead of simply plotting the average scores by cuisines, I chose to also show the full distribution of the scores, since averages are sensitive to extreme values.

Question 5

Is there a relationship between cuisine type and violation? For instance, do some cuisine types tend to have more of some type of violations that other cuisine types?

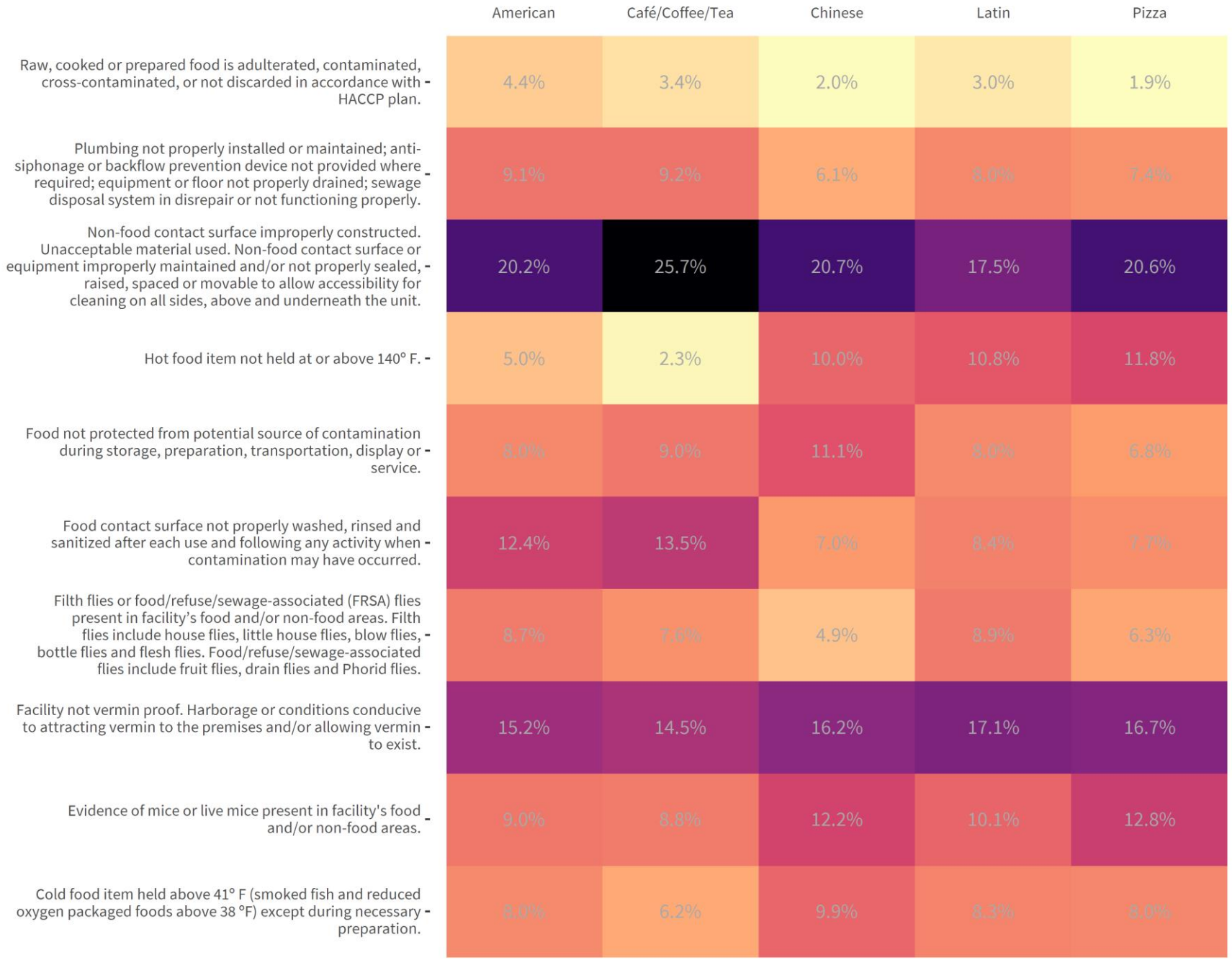
Most frequent cuisine type for each violation

Considering the top 5 cuisines and the top 10 violations. Horizontal values (by violation type) add up to 100%.



Most frequent violations for each cuisine type

Considering the top 5 cuisines and the top 10 violations. Vertical values (by cuisine type) add up to 100%.



Comments

Here, I first plotted, for each infraction, the cuisine types that incurred in more infractions. But since “American” cuisine is the most prevalent cuisine in the dataset, the plot would be somewhat unbiased.

To fix that, I decided to plot a chart showing the most frequent violation for each cuisine type. The difference is subtle, but the resulting plot is very different from the first one (and more interesting).