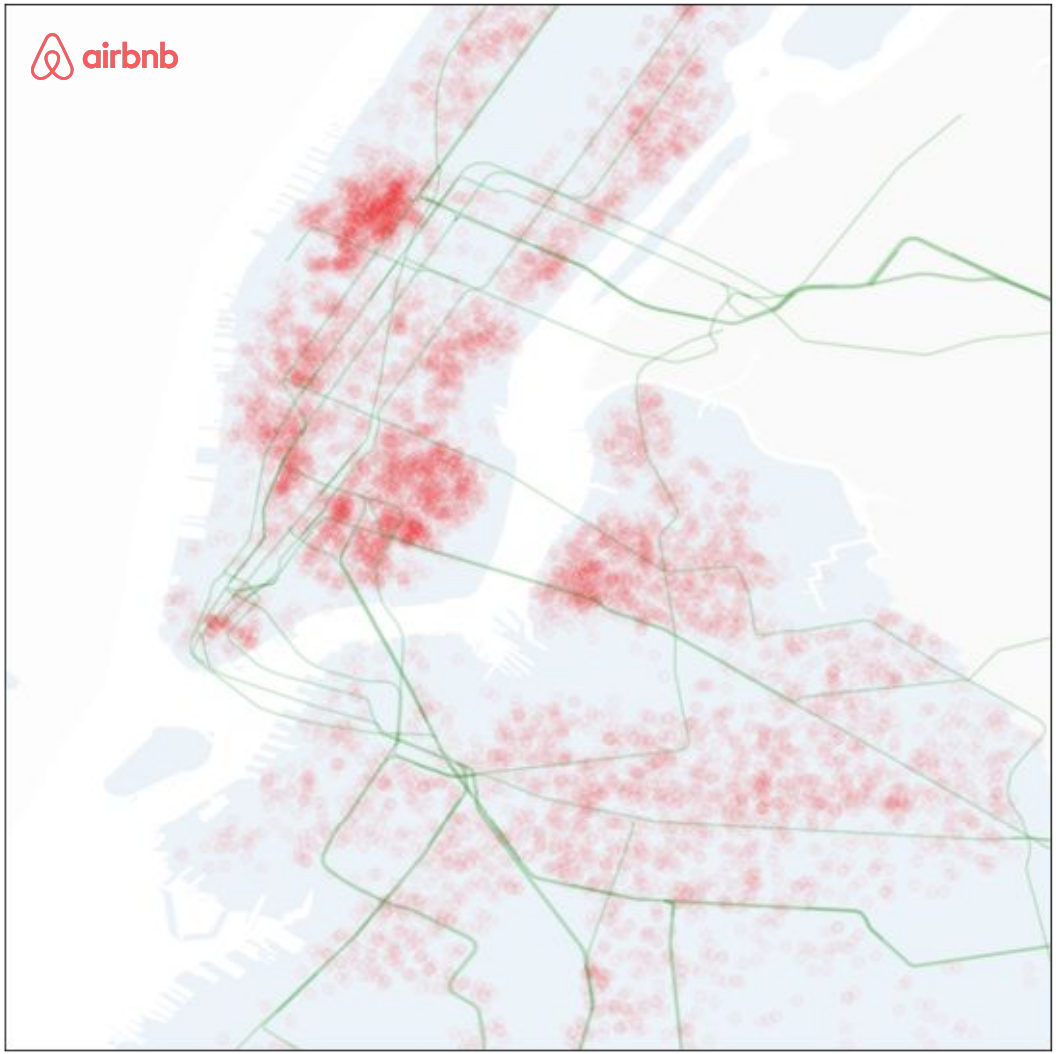
# A New Yorker’s Guide to Airbnb Pricing

Dan Segal
@djsegal

New York is a city that’s intimately connected to its train system

# Can you use train data to price out Airbnb rentals?



# Working with Airbnb Data

# Airbnb provides monthly data for major cities all the way back to 2015

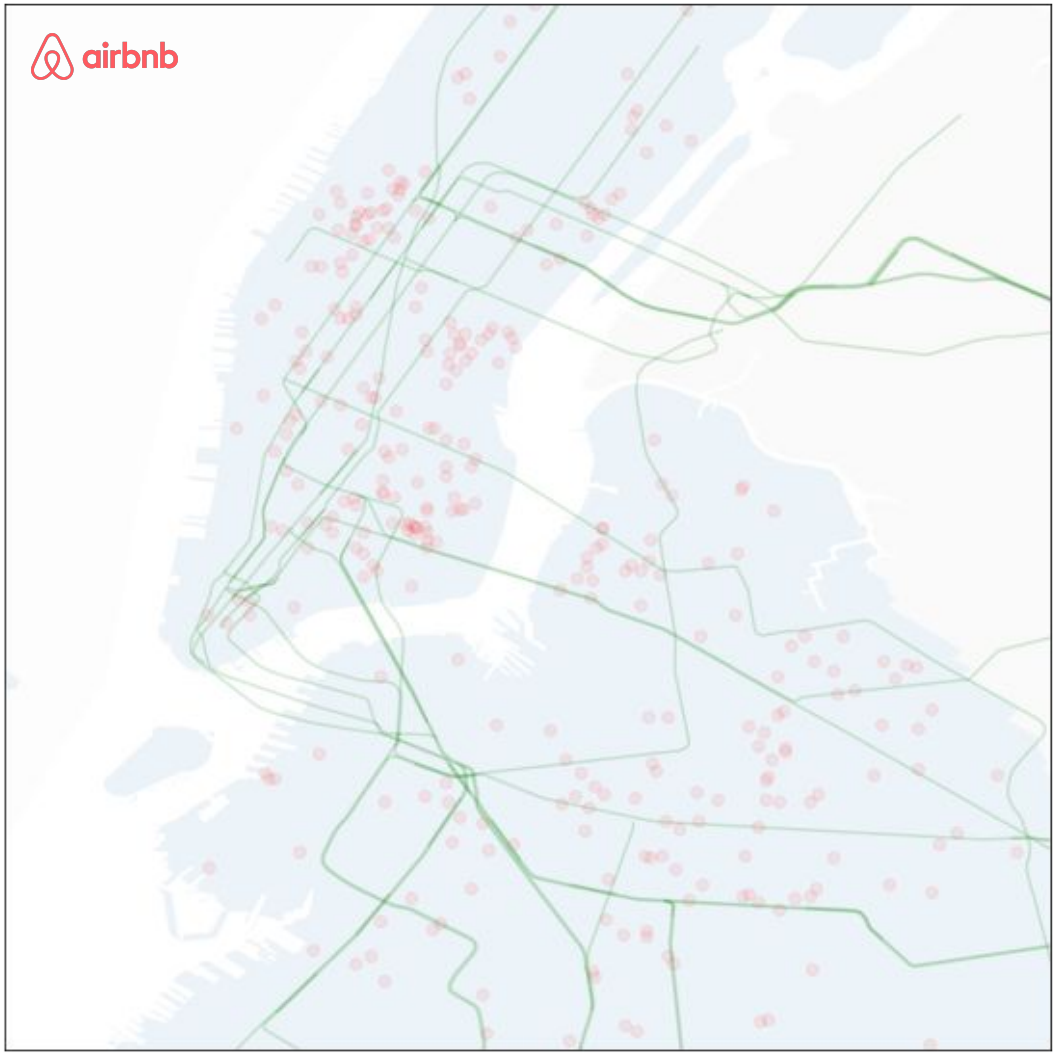
**Mutables**

- price
- rating
- reviews
- accommodates
- bathrooms
- bedrooms
- beds

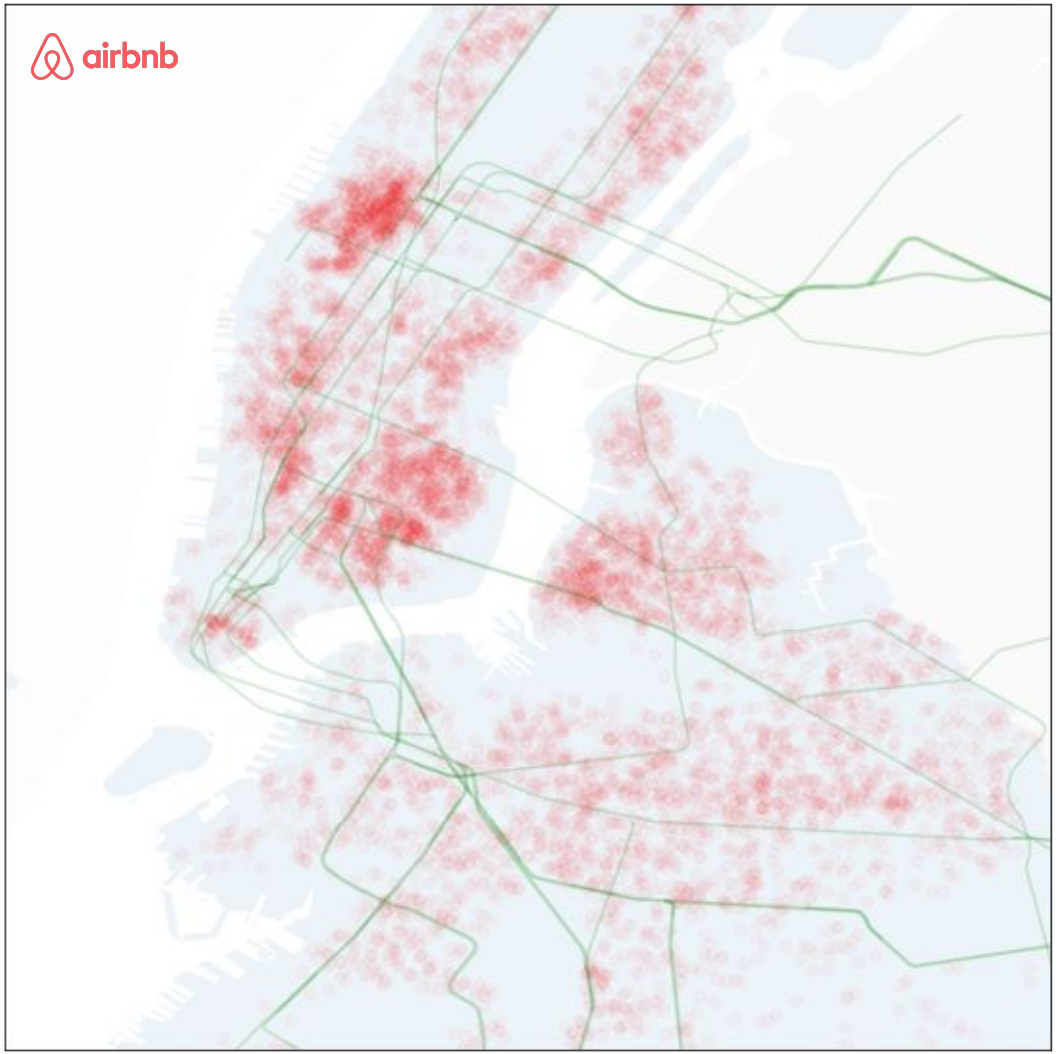
**Immutables**

- borough
- neighbourhood
- property type
- month
- year
- latitude
- longitude

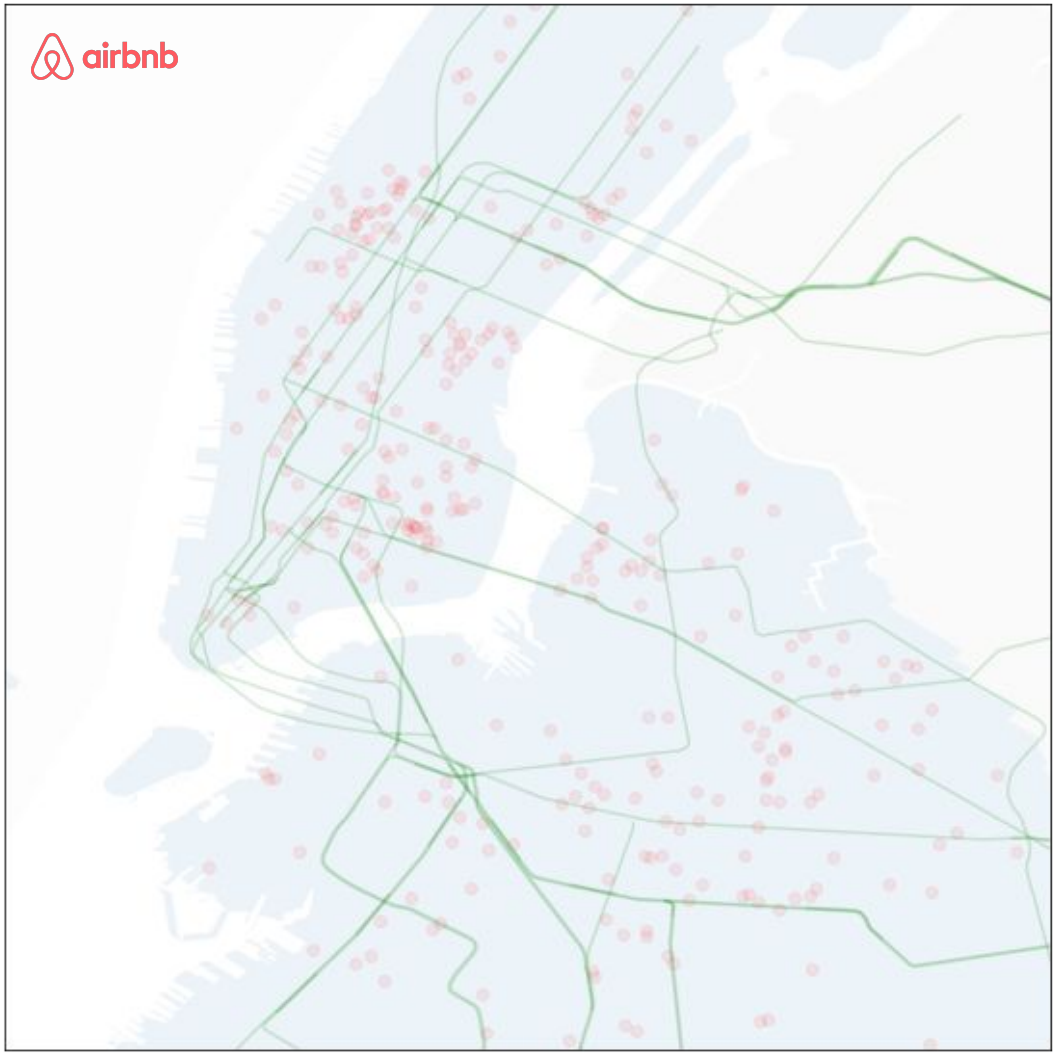
Using just Airbnb data, we get an $R^2$ value of ~39%



Using just Airbnb data, we get an $R^2$ value of ~39%



Using just Airbnb data, we get an $R^2$ value of ~39%



# Adding MTA Data

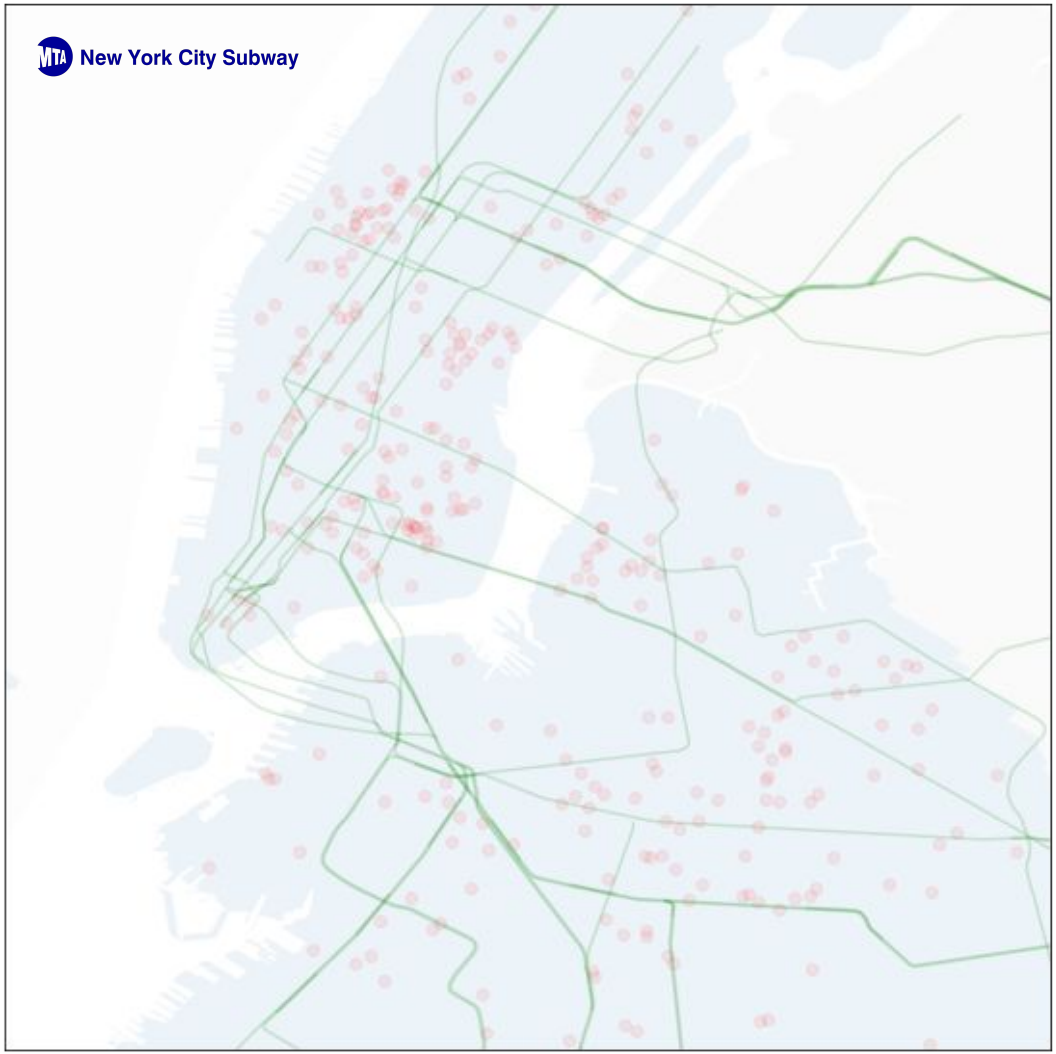
## The goal now is to incorporate rental locations into price

This includes finding:

- the distance to the nearest train
- the number of train lines nearby

# Adding train data improves $R^2$ by 4% (up to 43%)

# Adding train data improves $R^2$ by 4% (up to 43%)



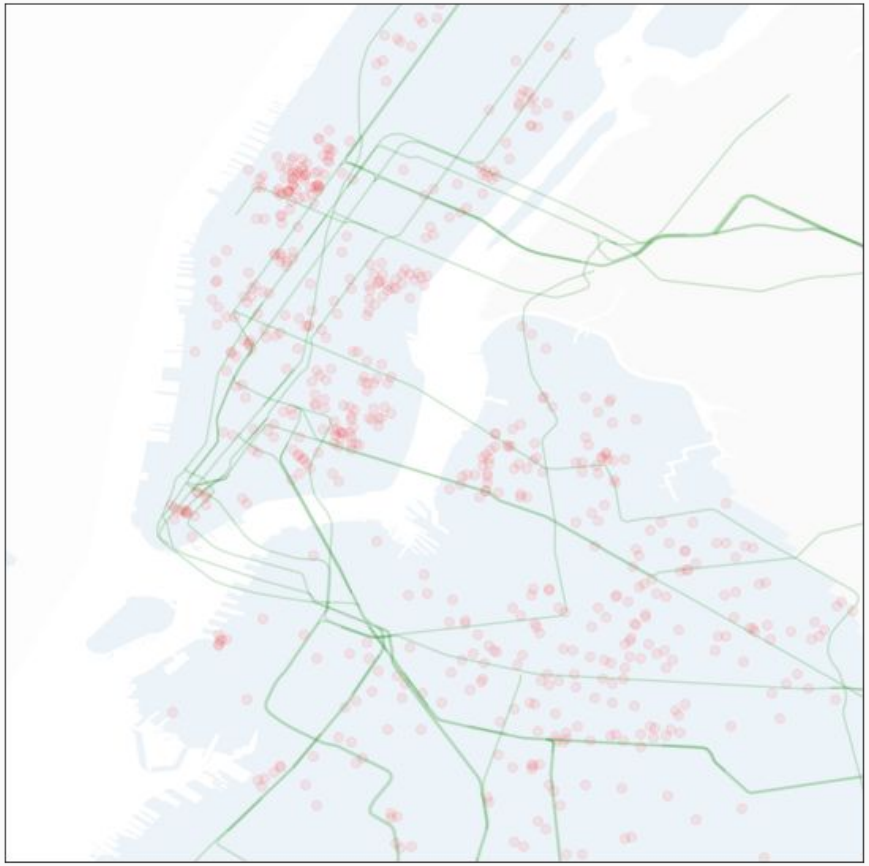
# Adding train data improves $R^2$ by 4% (up to 43%)

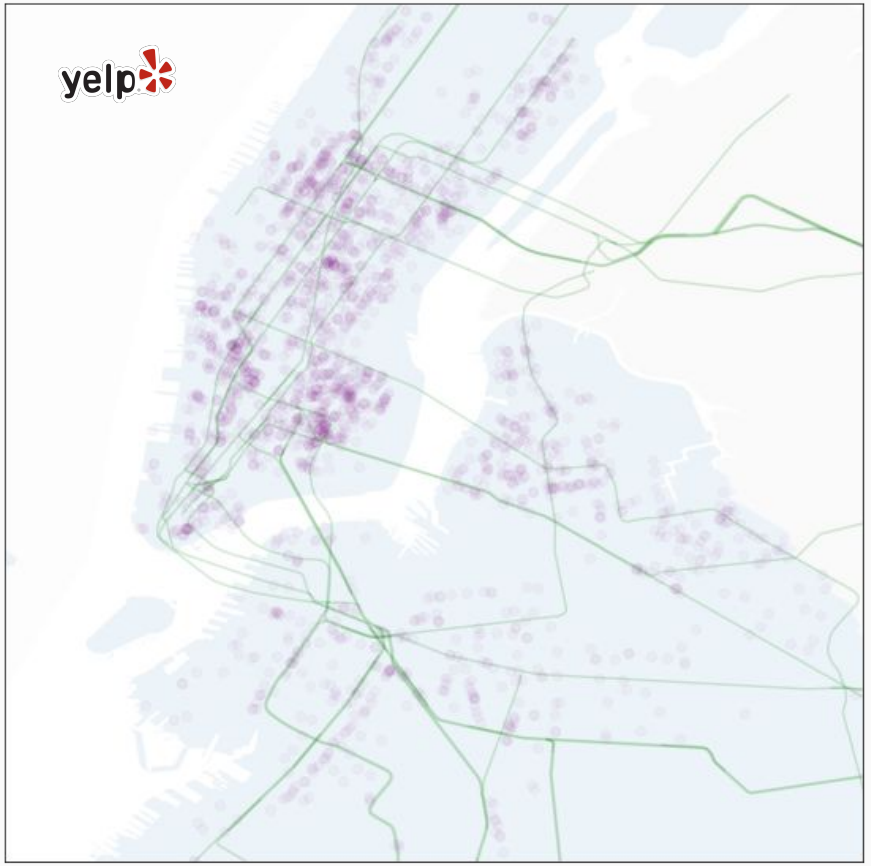
The question now is how do you factor in livability

# In a given day, a tourist may want to go to a...

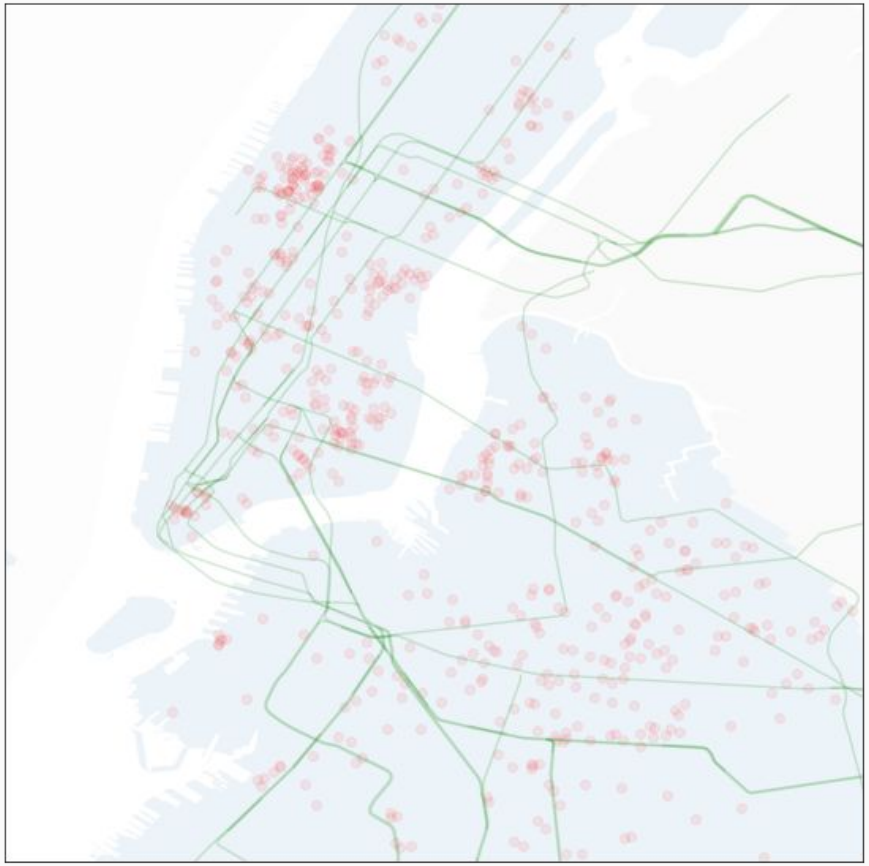
- **Bar (Yelp)**
- **Cafe (DOH Inspections)**
- **Restaurant (DOH Inspections)**
- Landmark (Wikipedia)
- Bodega (NYS Food Stores)

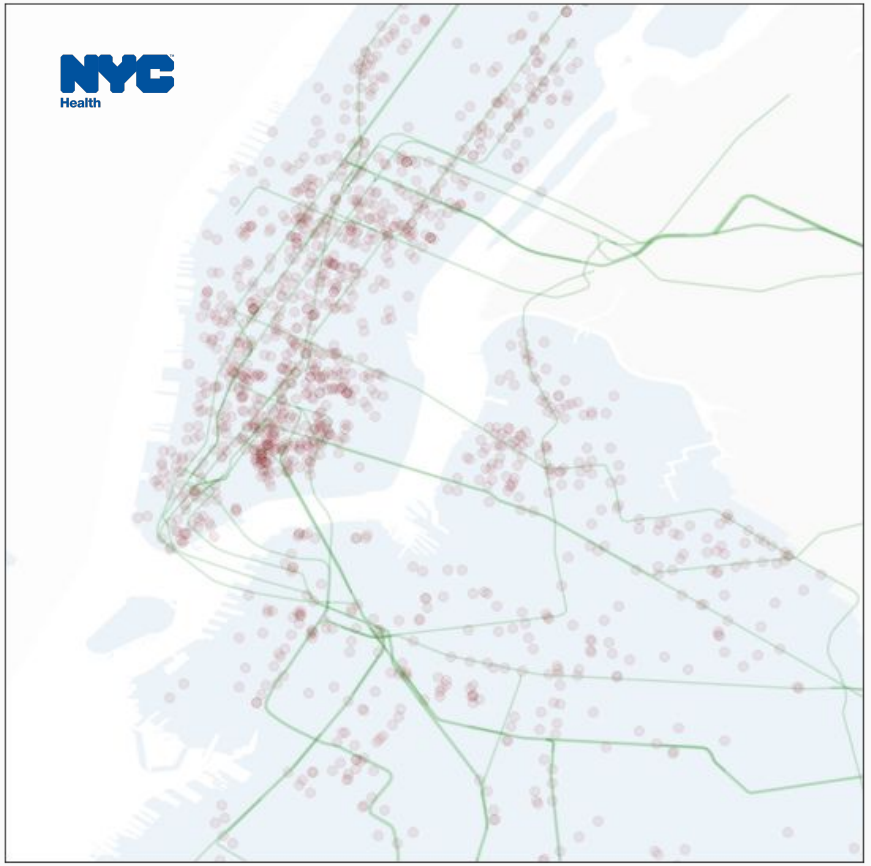
# Where are the **bars** in NYC?



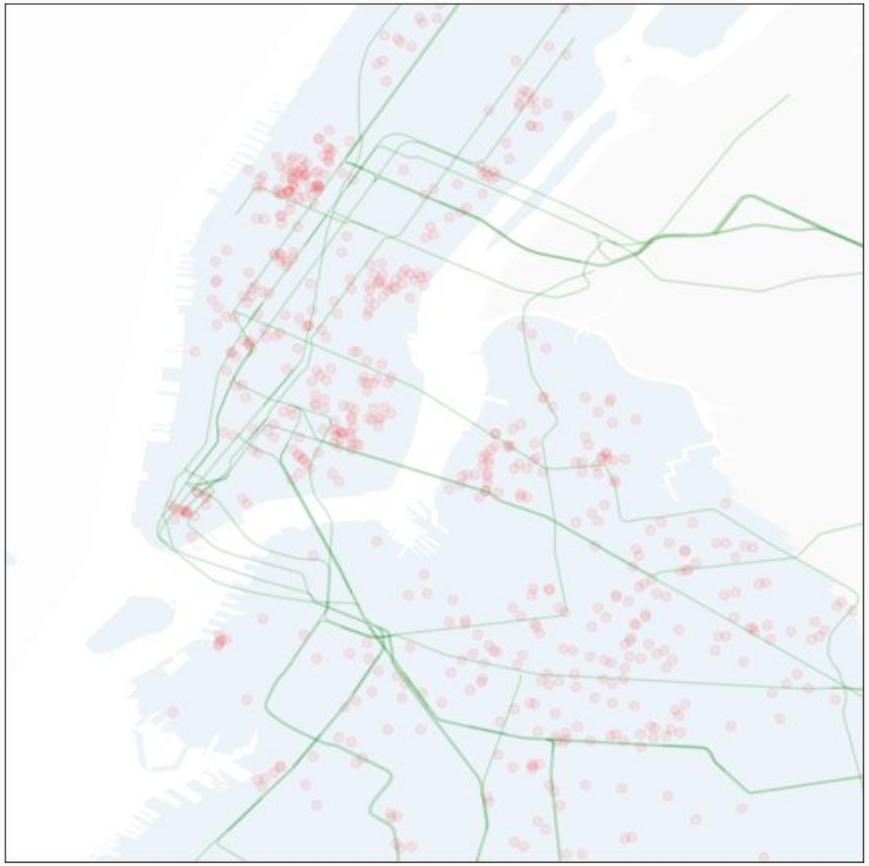


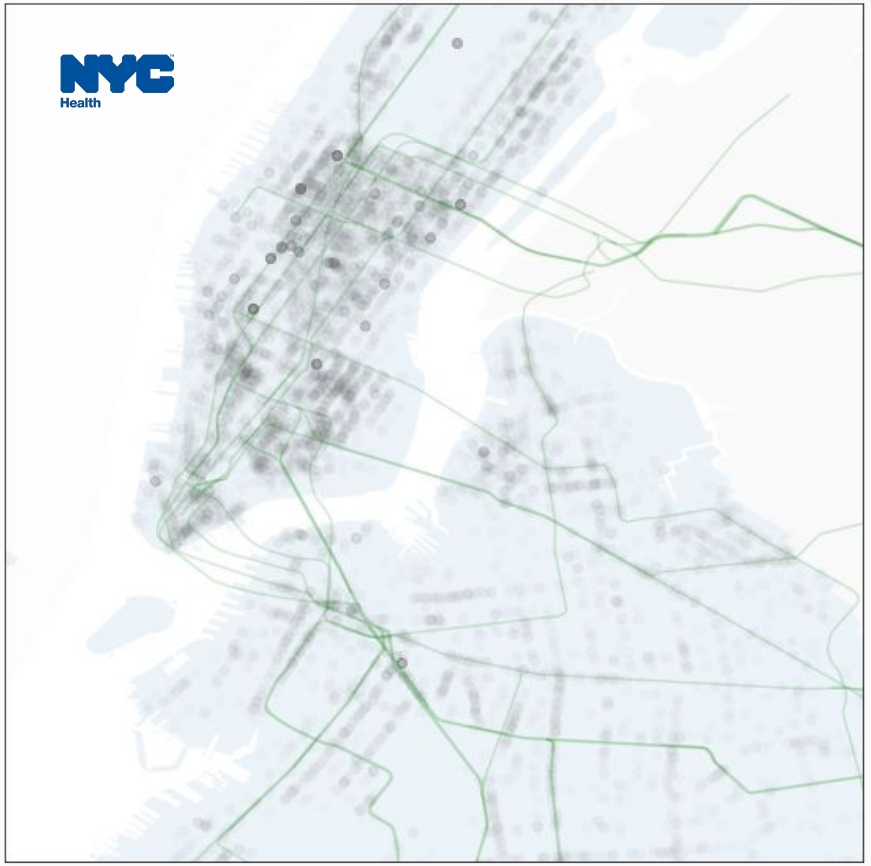
# Where are the **cafes** in NYC?



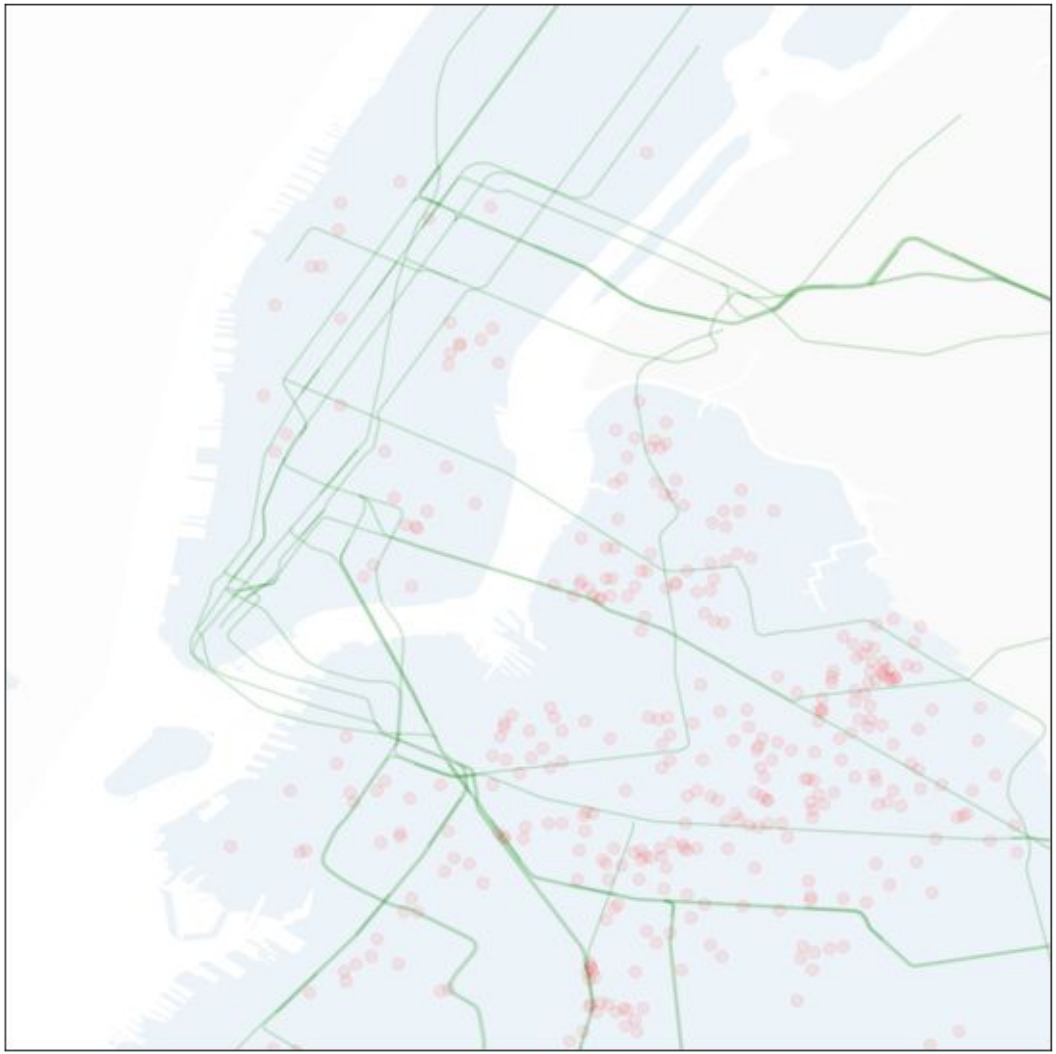


# Where are the **restaurants** in NYC?





Using bars, cafes, and restaurants bumps the $R^2$ up by 8% (to 51%)!



# Moving forward, these steps could be added to the model:

1. Factor in distance to landmarks (and the city)
2. Switch distances from metric to time
3. Price houses using valuations, grocers, and schools

# Appendix

# List of Coefficients for: Airbnb Only

| Column Name | Coefficient Value |
|---|---|
| is_brooklyn | -0.0781 |
| bathrooms | +0.0462 |
| accommodates | +0.0391 |
| bedrooms | +0.0382 |
| is_loft | +0.0292 |
| years_since_2015 | -0.0160 |
| accommodates_squared | -0.0145 |

# List of Coefficients for: With Trains

| Column Name | Coefficient Value |
|---|---|
| is_brooklyn | -0.1625 |
| is_brooklyn_norm_unique_lines | -0.0650 |
| is_brooklyn_log_trains_count | +0.0528 |
| bathrooms | +0.0459 |
| accommodates | +0.0432 |
| bedrooms | +0.0387 |
| norm_unique_lines | +0.0311 |

# List of Coefficients for: With Livability

| Column Name | Coefficient Value |
|---|---|
| is_brooklyn_norm_unique_food | +0.2226 |
| is_brooklyn_log_food_count | +0.1595 |
| accommodates | +0.0529 |
| is_brooklyn_log_bars_count | +0.0492 |
| bedrooms | +0.0450 |
| norm_unique_food | -0.0402 |
| bathrooms | +0.0393 |