Lab 3

Elizabeth Walter

19 September 2021

Essential pandemic construction

The New York City Department of Buildings (DOB) maintains a list of construction sites that have been categorized as "essential" during the city's shelter-in-place pandemic order.

Load and clean data

First we load and clean the data.

```
# You'll only need the tidyverse library for this exercise
library(tidyverse)

# Load original data; note that this needs to point to the file you downloaded.
essential_raw <- read_csv("EssentialConstruction.csv")

# Clean the data a little
# Some of the borough names are in ALL CAPS, so we use str_to_title() to convert
# everything in the column to title case.
# We also make BOROUGH and CATEGORY factors (or categorical variables)
essential <- essential_raw %>%
  mutate(BOROUGH = str_to_title(BOROUGH),
         BOROUGH = factor(BOROUGH),
         CATEGORY = factor(CATEGORY))
```
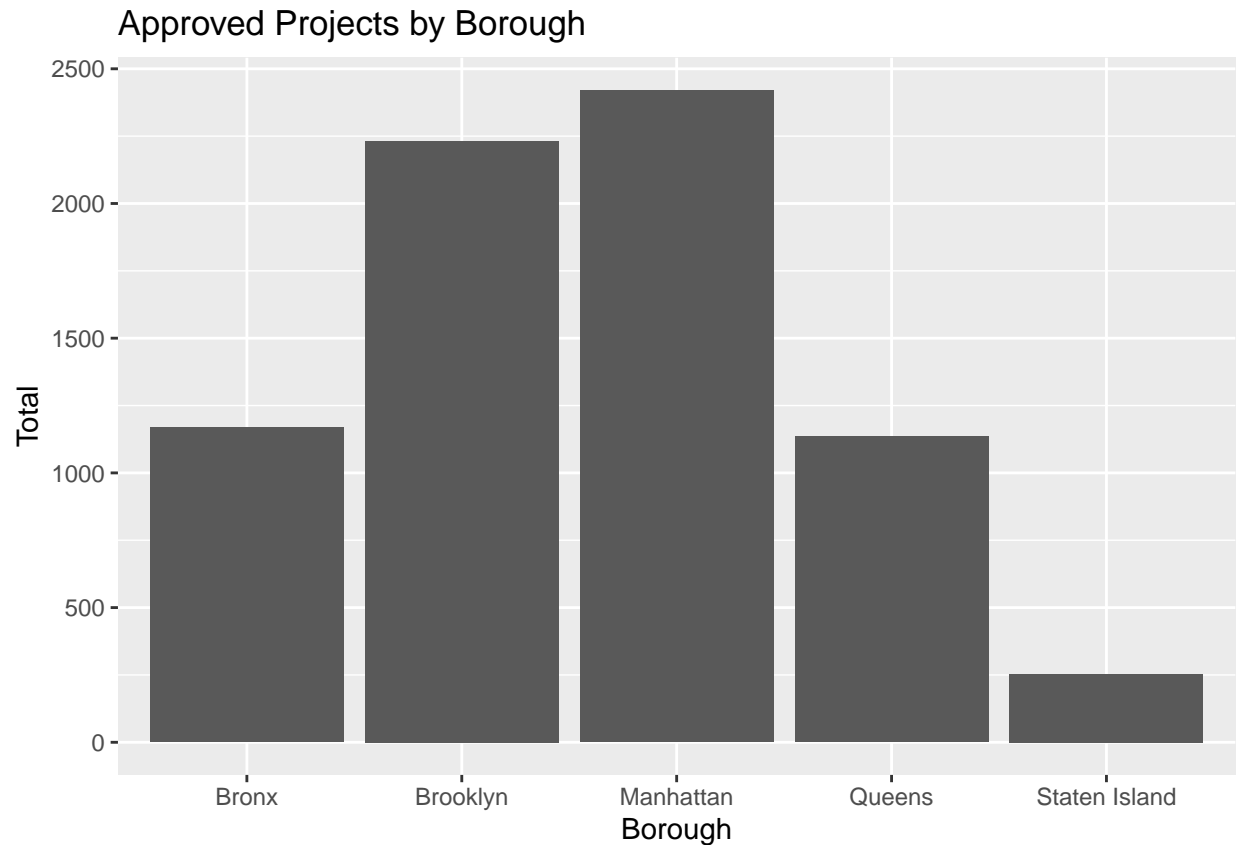
Approved projects by borough

Right now there's a row for each approved construction site. We need to condense that down to get counts of construction sites by different variables. We can do this by using `group_by()` and `summarize()`

```
essential_by_borough <- essential %>%
  group_by(BOROUGH) %>%
  summarize(total = n()) %>%
  mutate(proportion = total / sum(total))
```
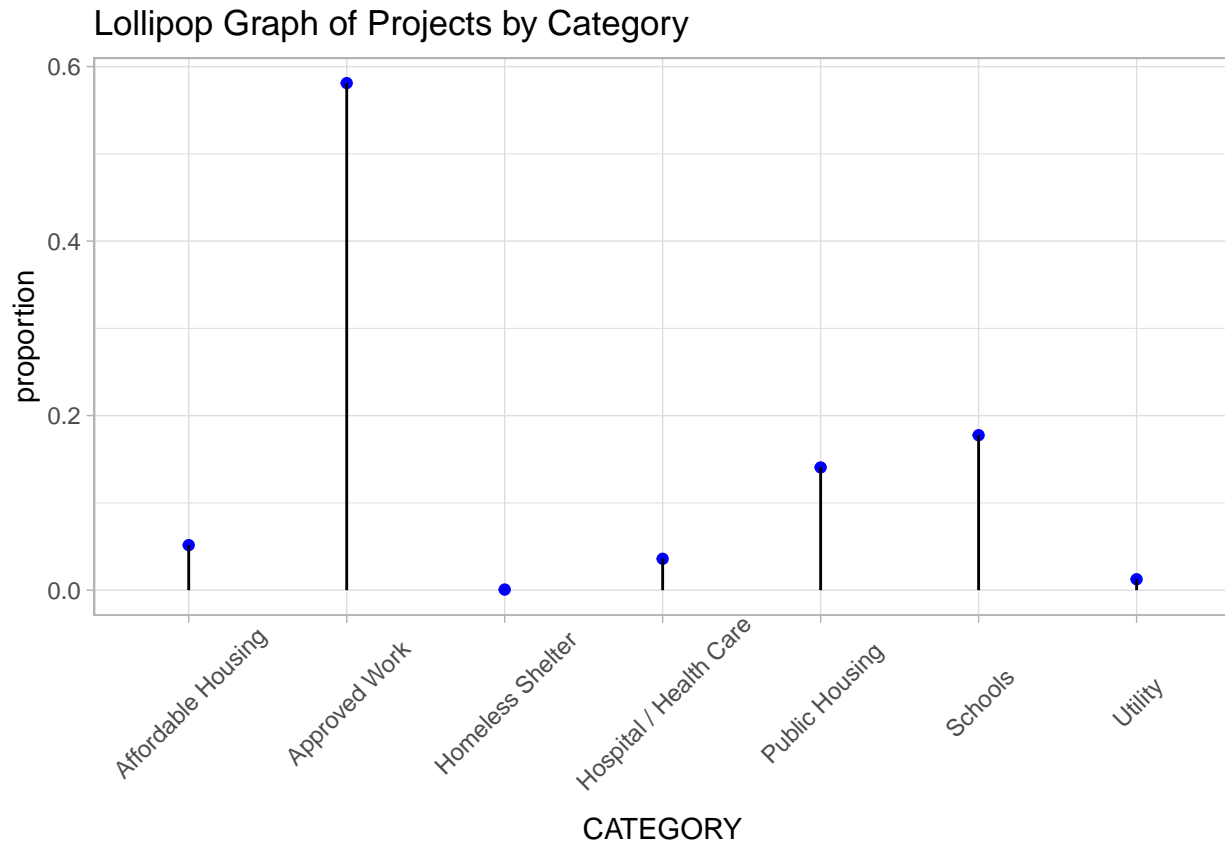
```
# Add plot with geom_col() here
apprvd_by_bor <- ggplot(data = essential_by_borough, aes(x = BOROUGH, y = total)) +
  geom_col() +
  labs(x = 'Borough', y= 'Total',
       title = 'Approved Projects by Borough')
apprvd_by_bor
```

Approved Projects by Borough

Approved projects by category

```r
# Create a summarized dataset of projects by category
essential_by_categ <- essential %>%
  group_by(CATEGORY) %>%
  summarize(total = n()) %>%
  mutate(proportion = total / sum(total))
```

```r
# Add a lollipop chart here
apprvd_by_categ <- ggplot(data = essential_by_categ, aes(x = CATEGORY, y = proportion)) +
  geom_point(color = 'blue') +
  geom_segment(aes(x= CATEGORY, xend = CATEGORY, y = 0, yend = proportion)) +
  labs(title = "Lollipop Graph of Projects by Category") +
  theme_light() +
  theme(axis.text.x = element_text(angle = 45, vjust = .6))
apprvd_by_categ
```

Lollipop Graph of Projects by Category

Info/code for lollipop graphs from: https://www.r-graph-gallery.com/300-basic-lollipop-plot.html and https://www.r-graph-gallery.com/301-custom-lollipop-chart.html#general

Approved projects across borough and category

```
# Create a summarized dataset of projects by both borough and category

essential_bor_categ <- essential %>%
  select(BOROUGH, CATEGORY) %>%
  group_by(BOROUGH, CATEGORY) %>%
  summarize(tot_cat= n()) %>%
  group_by(BOROUGH) %>%
  mutate(tot_bor = sum(tot_cat),
         cat_prop_of_bor = tot_cat / tot_bor)
```
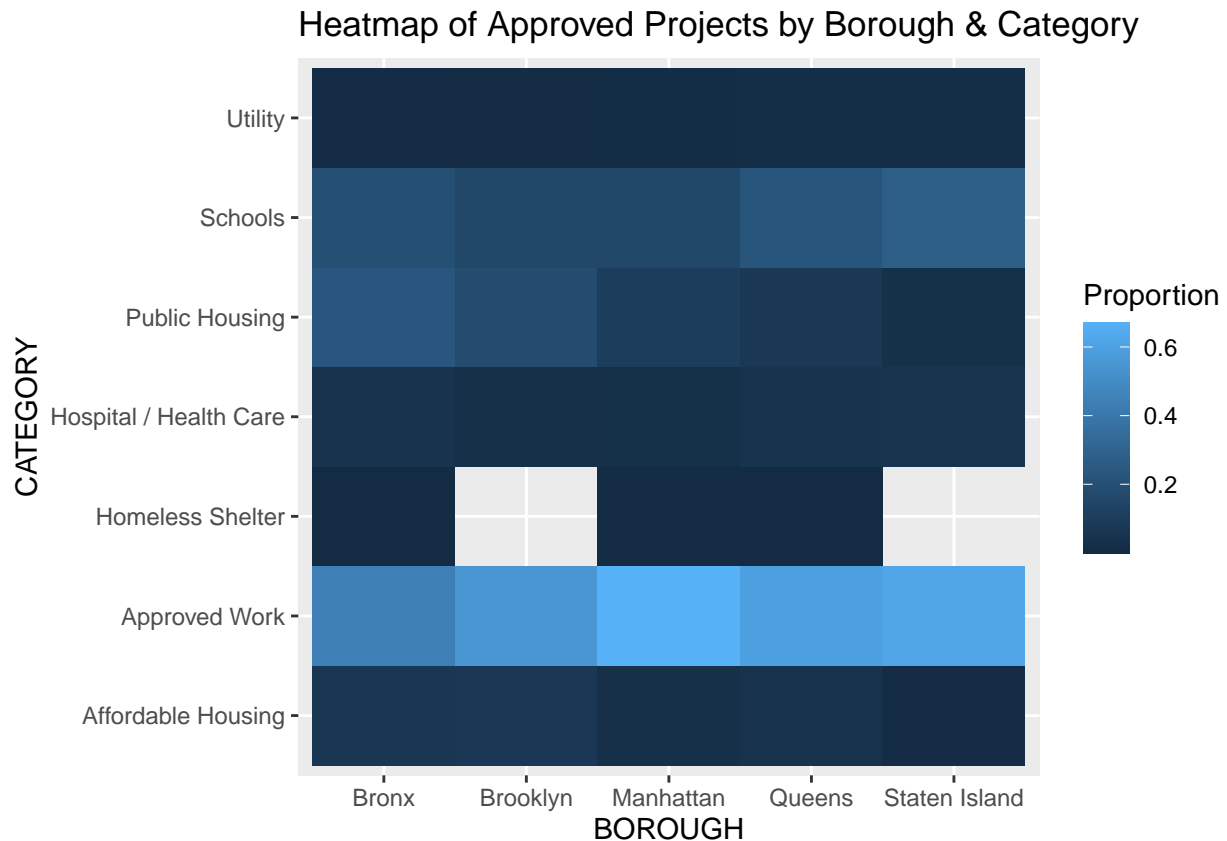
`summarise()` has grouped output by 'BOROUGH'. You can override using the `.groups` argument.

```
essential_bor_categ
```

```
## # A tibble: 33 x 5
## # Groups:   BOROUGH [5]
##    BOROUGH  CATEGORY              tot_cat tot_bor cat_prop_of_bor
##    <fct>    <fct>                   <int>   <int>           <dbl>
##  1 Bronx    Affordable Housing         80    1169          0.0684
```

```
## 2 Bronx   Approved Work              518   1169      0.443
## 3 Bronx   Homeless Shelter             1   1169      0.000855
## 4 Bronx   Hospital / Health Care      55   1169      0.0470
## 5 Bronx   Public Housing             276   1169      0.236
## 6 Bronx   Schools                    229   1169      0.196
## 7 Bronx   Utility                     10   1169      0.00855
## 8 Brooklyn Affordable Housing        168   2231      0.0753
## 9 Brooklyn Approved Work            1223   2231      0.548
## 10 Brooklyn Hospital / Health Care    66   2231      0.0296
## # ... with 23 more rows
```

```
# Add a heatmap here with geom_tile()
bor_categ_heatmap <- ggplot(data = essential_bor_categ, aes(x = BOROUGH, y = CATEGORY,
                                                            fill = cat_prop_of_bor)) +
  geom_tile() +
  guides(fill = guide_colorbar(title = "Proportion")) +
  labs(title = "Heatmap of Approved Projects by Borough & Category")
bor_categ_heatmap
```


Additional information/code about heatmaps from: https://r-charts.com/correlation/heat-map-ggplot2/