Assignment3

Weijie Jaycee Xin

```
library(ggplot2)
library(readxl)
setwd('/Users/admin/Desktop/DataViz')
```

Plot 1

```
g1 <- read_excel('G2.xlsx')

ggplot(data=g1, aes(x=no_hs_tract, y=health_insurance_uninsured_tract, color=GEO)) +
  scale_colour_brewer(palette = "Set2") +
  geom_point(alpha=0.4) +
  geom_smooth(method='lm', se=FALSE)+
  scale_x_continuous(breaks=c(10, 20, 30, 40, 50, 60),
                     labels=c("10%","20%", "30%", "40%", "50%","60%")) +
  scale_y_continuous(breaks=c(0, 10, 20, 30, 40),
                     labels=c('0%', "10%", "20%", "30%", "40%")) +

  labs(y = "No Health Insurance Rate (% of total population)", x = "High School Drop-
out Rate (% of adults 25+)",
       title="Health Insurance V.S Education Level",
       subtitle="Examining the relationship between a lack of health insurance and hi
gh school dropout \non census tract level",
       caption="DATA2GO.NYC") +

  theme(legend.position = "bottom",
        plot.title = element_text( face="bold", size=15),
        plot.subtitle = element_text(size=10, margin=margin(b=10)),
        plot.caption = element_text(size=8, margin=margin(t=10)),
        plot.margin=unit(c(1,1,1,1),"cm"),
        panel.background = element_rect(fill = "#e9e9e9"),
        axis.text = element_text(size=8),
        axis.text.x = element_text(margin=margin(t=3)),
        axis.title = element_text(size=10),
        axis.title.x = element_text(margin=margin(t=10)),
        axis.title.y = element_text(margin=margin(r=10))) +
  guides(color=guide_legend("Borough"))
```

Health Insurance V.S Education Level

Examining the relationship between a lack of health insurance and high school dropout
on census tract level

High School Drop-out Rate (% of adults 25+)

Borough — Bronx — Kings — New York — Queens — Richmond
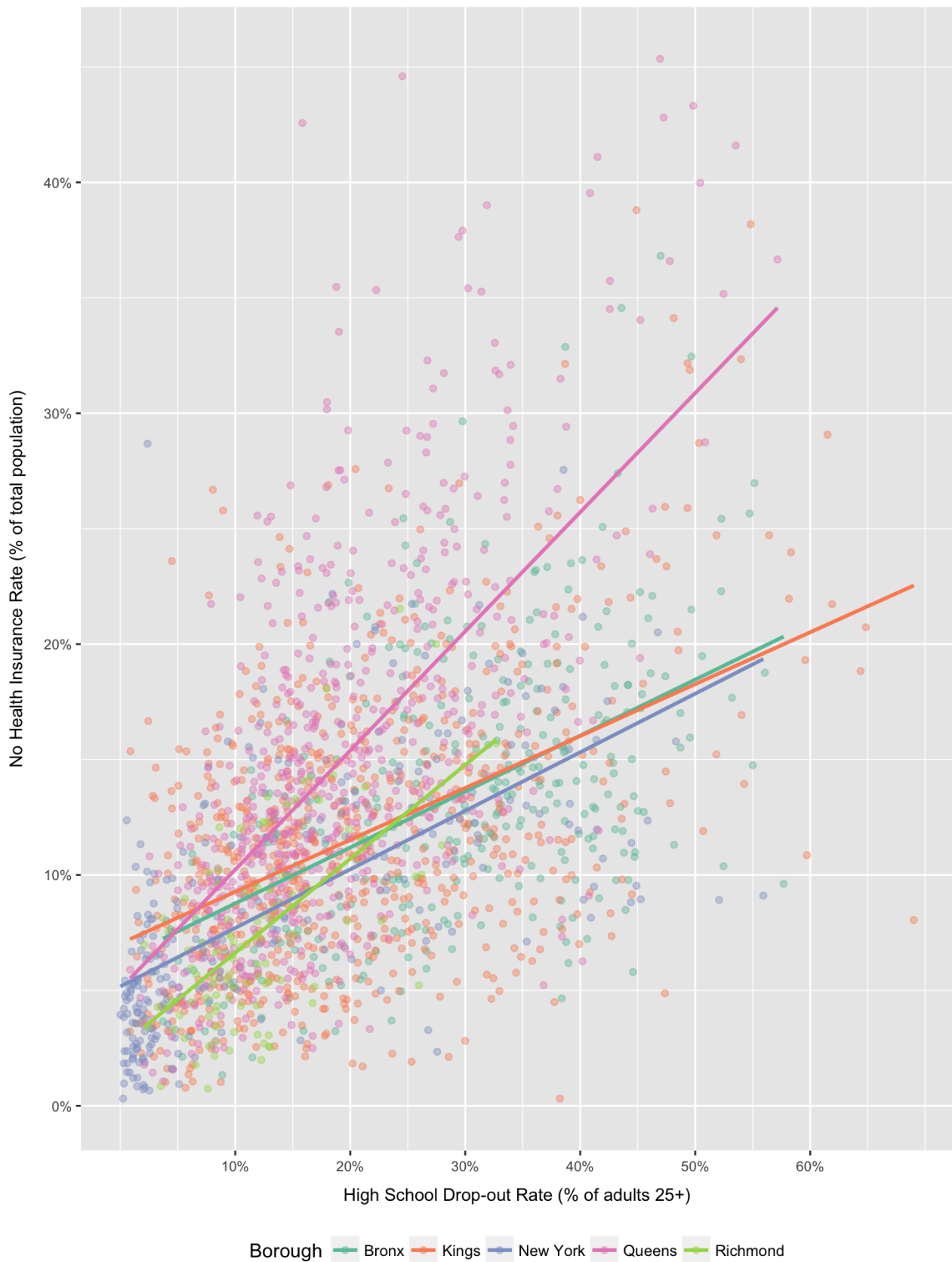
DATA2GO.NYC

Summary The graph is designed to examine the correlation between the social determinant variable *High School Drop-out Rate(2010-2014)* and the health outcome variable *No Health Insurance Rate(2011-2015)*. The data are collected on the census tract level of NYC, with over 2,000 data points. From the graph we can see,

there is a positive relationship between the two variables among all five boroughs. Bronx, Kings, New York and Richmond have similar slope rates, while Queens has a much steeper one.

Plot 2

```
g1 <- read_excel('G1.xlsx')

ggplot(data=g1, aes(x=Borough, y = alzheimers_deaths_cd, size=No_Exercise_Rate)) +
  geom_point(alpha=0.5, color='darkgoldenrod1') +
  scale_y_continuous(breaks=c(0, 2.5, 5, 7.5, 10, 12.5, 15, 17.5),
                     labels=c('0', "2.5", "5", "7.5", "10", '12.5', '15', '17.5')) +

  labs(y = "Alzheimer's Disease Deaths (per 100,000 residents)", x = "Boroughs",
       title="Alzheimer's Deaths V.S Exercise Level",
       subtitle="Examining the trend between Alzheimer's deaths rate and \npercentage
 of adults 18+ who did not exercise at all in the past 30 days \non NYC community dis
trict level",
       caption="DATA2GO.NYC") +

  theme(legend.position = "bottom",
        plot.title = element_text( face="bold", size=15),
        plot.subtitle = element_text(size=10, margin=margin(b=10)),
        plot.caption = element_text(size=8, margin=margin(t=10)),
        plot.margin=unit(c(1,1,1,1),"cm"),
        panel.background = element_rect(fill = "#e9e9e9"),
        axis.text = element_text( size=8),
        axis.text.x = element_text(angle=0, margin=margin(t=3)),
        axis.title = element_text( size=10),
        axis.title.x = element_text(margin=margin(t=10)),
        axis.title.y = element_text(margin=margin(r=10))) +

  guides(size=guide_legend("No Exercise Rate"))
```
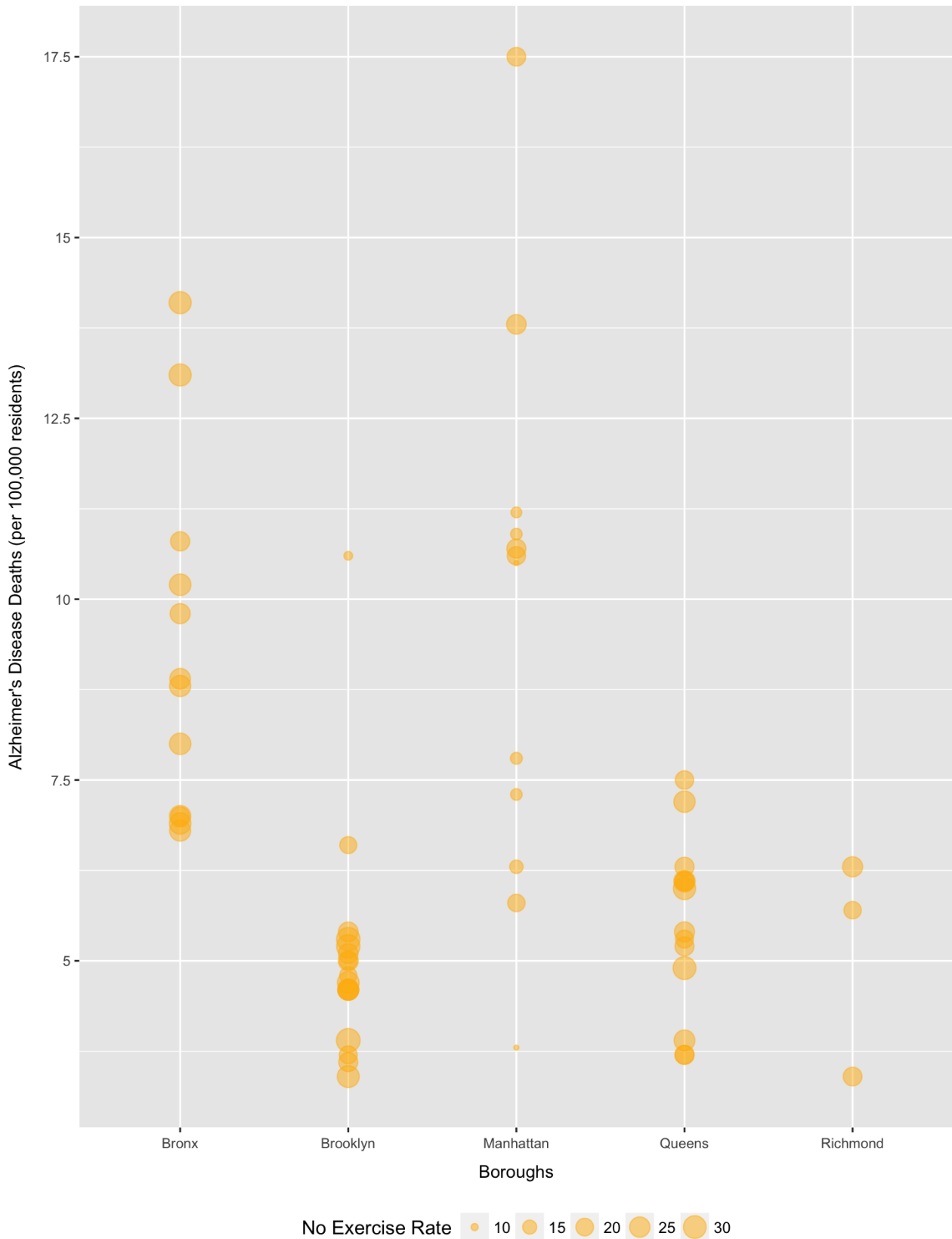
Alzheimer's Deaths V.S Exercise Level

Examining the trend between Alzheimer's deaths rate and
percentage of adults 18+ who did not exercise at all in the past 30 days
on NYC community district level

DATA2GO.NYC

Summary The graph is designed to examine the relationship between the social determinant variable *No Exercise Rate(2011-2013)*, which is percentage of adults 18+ who did not exercise at all in the past 30 days, and the health outcome variable *Alzheimer's Deaths per 100,000 residents(2009-2013)*. The data are collected

on the community district level of NYC, with 59 data points. From the graph we can see, generally, there is no strong relationship between the two variables among all five boroughs. In Bronx, Queens and Richmond, *No Exercise Rate* barely changes with the increase of *Alzheimer's Deaths*. In Brooklyn, there is a weak negative relationship between the two variables, while in Manhattan, there is a weak positive relationship, which is the relationship people may presume.

Plot 3

```
g4 <- read_excel('G4.xlsx')

ggplot(g4, aes(x=homeless_children_rate_cd, y=infant_mortality_cd)) +
  geom_point(size=0.8, alpha=0.5, aes(color=Borough)) + geom_line(size=0.7,
alpha=0.8, aes(color=Borough)) +
  scale_colour_brewer(palette = "Set2") + geom_smooth(method = 'loess', se=FALSE, col
or='black') +
  scale_x_continuous(breaks=c(0, 5, 10, 15, 20, 25, 30, 35),
                     labels=c('0', "5","10", "15", "20", "25","30",'35')) +
  scale_y_continuous(breaks=c(1.5, 3, 4.5, 6, 7.5, 9),
                     labels=c("1.5", "3", "4.5", "6", "7.5", '9')) +

  labs(y = "Infant Mortality (death before first birthday per 1,000 live births)", x
= "Homeless Children (per 1,000 children ages 0-17)",
       title="Infant Mortality V.S Homeless Children",
       subtitle="Examining the relationship between infant mortality and homeless chi
ldren rate \non NYC community district level",
       caption="NYC Environment & Health Data Portal") +

  theme(legend.position = "bottom",
        plot.title = element_text( face="bold", size=15),
        plot.subtitle = element_text(size=10, margin=margin(b=10)),
        plot.caption = element_text(size=8, margin=margin(t=10)),
        plot.margin=unit(c(1,1,1,1),"cm"),
        panel.background = element_rect(fill = "#e9e9e9"),
        axis.text = element_text(size=8),
        axis.text.x = element_text(margin=margin(t=3)),
        axis.title = element_text(size=10),
        axis.title.x = element_text(margin=margin(t=10)),
        axis.title.y = element_text(margin=margin(r=10)))
```

Infant Mortality V.S Homeless Children

Examining the relationship between infant mortality and homeless children rate
on NYC community district level
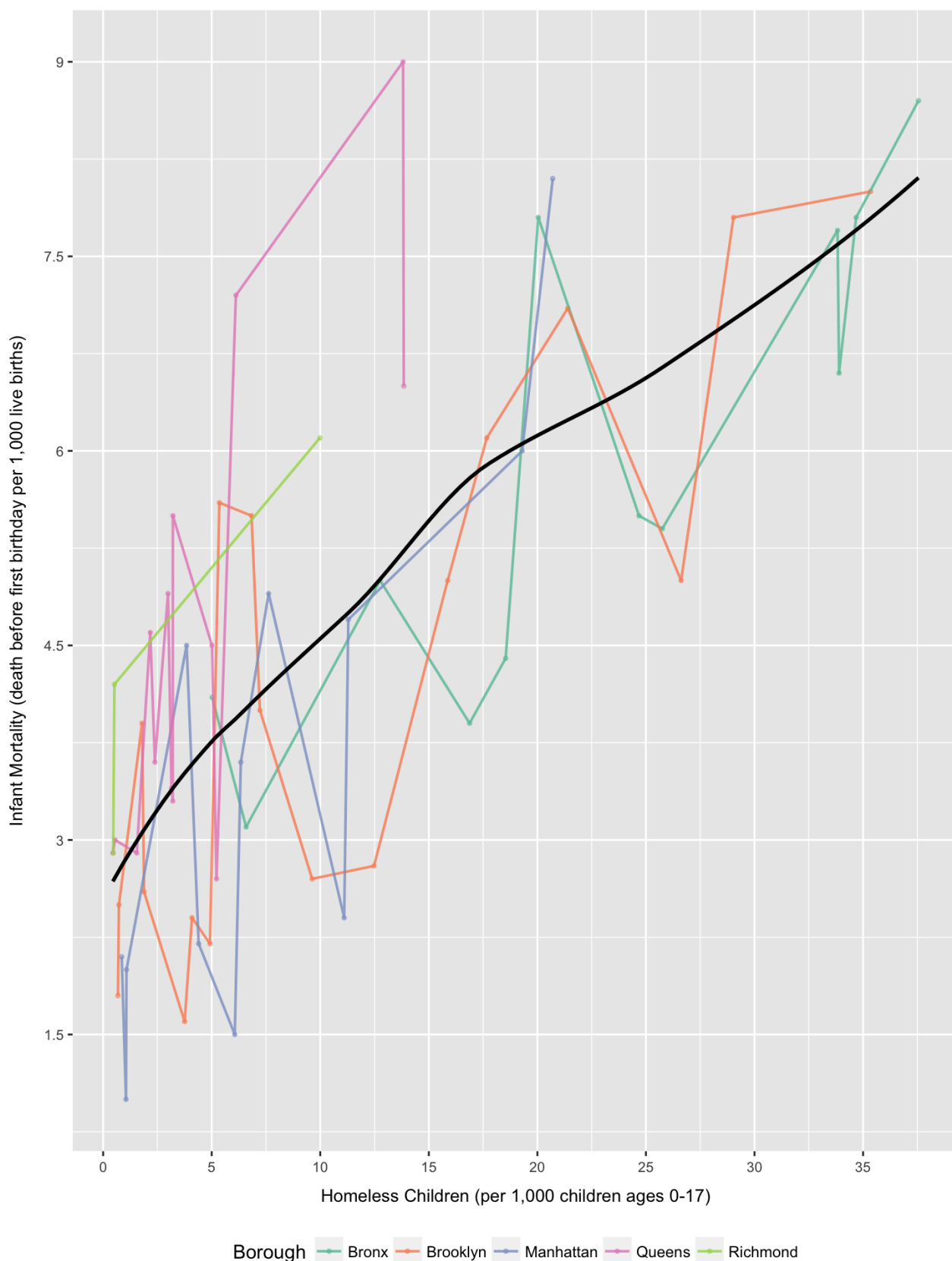
Summary The graph is designed to examine the relationship between the social determinant variable *Homeless Children(2015)* and the health outcome variable *Infant Mortality(2011-2013)*. The data are collected on the community district level of NYC, with 59 data points. From the graph we can see, there is a positive

relationship between the two variables among all five boroughs. Bronx and Brooklyn have clearer and stronger trends compared to the other three boroughs.

Plot 4

```
g1 <- read_excel('G1.xlsx')

ggplot(data=g1, aes(x=diabetes, y = GEO_DISPLAY_NAME, color=Borough)) +
  scale_colour_brewer(palette = "Set2") +
  geom_segment(aes(yend=GEO_DISPLAY_NAME,xend=0)) +
  geom_point() +
  facet_grid(Borough ~ ., scales="free_y", space="free_y") +
  scale_x_continuous(breaks=c(0, 2.5, 5, 7.5, 10, 12.5, 15, 17.5),
                     labels=c('0%', "2.5%", "5%", "7.5%", "10%", '12.5%', '15%', '17.
5%')) +

  labs(y = "Community District", x = "Diabetes Diagnosis Rate (% of adults 18+)",
       title="Diabetes Diagnosis Rate Description",
       subtitle="Describing diabetes diagnosis rate on NYC community district level",
       caption="DATA2GO.NYC") +

  theme(legend.position = "bottom",
        legend.title = element_text(size=10),
        legend.text = element_text(size=8),
        panel.grid.major.y=element_blank(),
        panel.grid.major.x = element_line(color='white', size=0.25),
        plot.title = element_text( face="bold", size=15),
        plot.subtitle = element_text(size=10, margin=margin(b=10)),
        plot.caption = element_text(size=8, margin=margin(t=10)),
        plot.margin=unit(c(1,1,1,1),"cm"),
        panel.background = element_rect(fill = "#e9e9e9"),
        axis.text.y = element_text(size=7),
        axis.text.x = element_text(angle=90, hjust=1, margin=margin(t=-1), size=8),
        axis.title = element_text(size=10),
        axis.title.x = element_text(margin=margin(t=10)),
        axis.title.y = element_text(margin=margin(r=10)),
        strip.text.y = element_text(angle=0, margin=margin(l=0)),
        axis.ticks=element_blank())
```
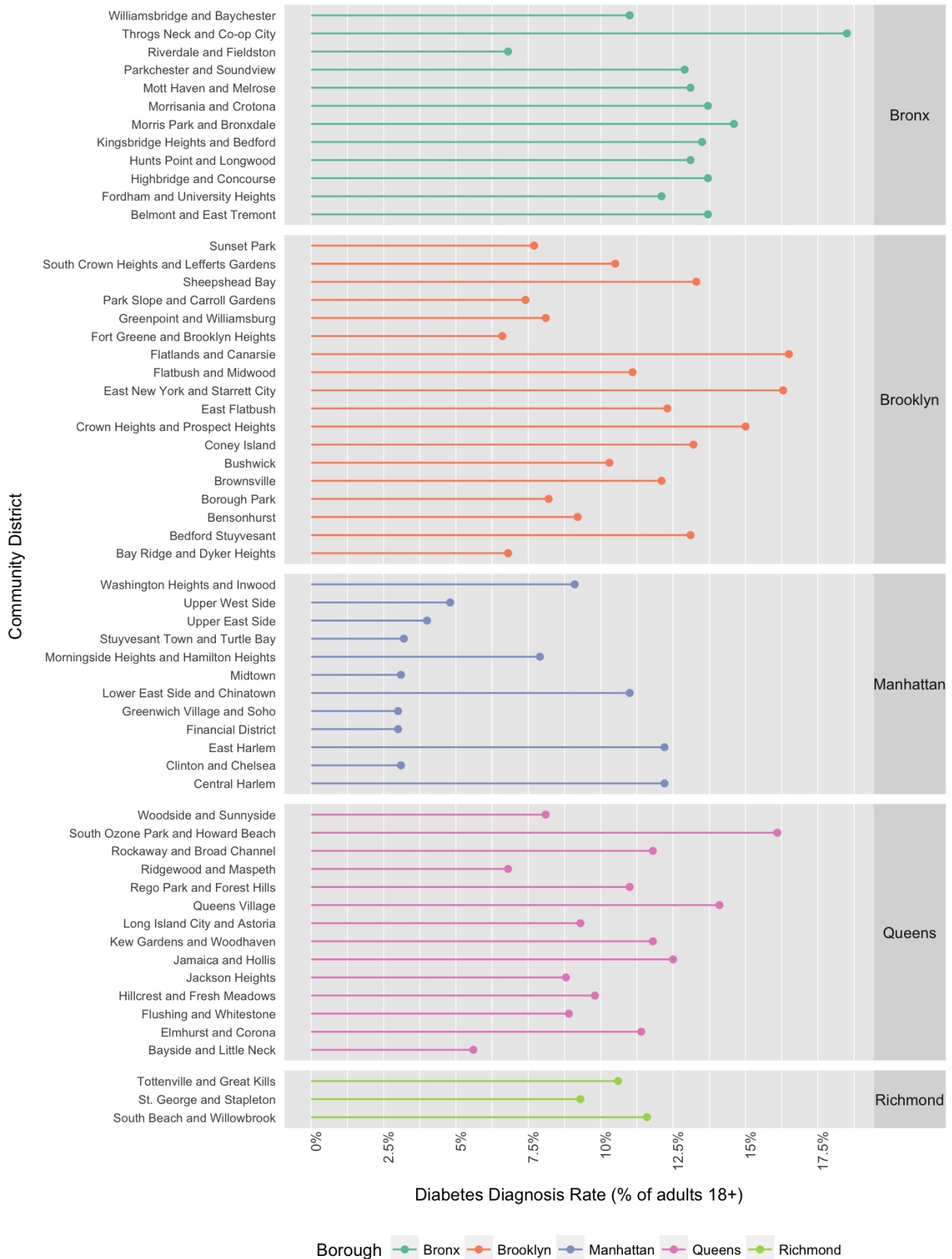
Diabetes Diagnosis Rate Description

Describing diabetes diagnosis rate on NYC community district level

Diabetes Diagnosis Rate (% of adults 18+)

Borough ● Bronx ● Brooklyn ● Manhattan ● Queens ● Richmond

DATA2GO.NYC

Summary The graph is designed to examine the distribution of *Diabetes Diagnosis Rate(2011-2013)*, which is the percentage of adults 18+ who have ever been told they have diabetes. The data are collected on the community district level of NYC, with 59 data points. From the graph we can see, Bronx, Brooklyn, Queens

and Richmond have similar distributions, while Manhattan has an overall lower rate distribution.

Plot 5

```
g4 <- read_excel('G4.xlsx')

ggplot(data=g4, aes(x=cancer_deaths_cd, fill=Borough)) +
  scale_colour_brewer(palette = "Set2") +
  geom_density(alpha=0.4) +
  scale_x_continuous(breaks=c(100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200),
                     labels=c('100', '110', "120", "130", "140", "150", '160', '170',
 '180', '190', '200')) +

  labs(y = "Density", x = "Cancer Deaths (per 100,000 residents)",
       title="Cancer Deaths Rate Distribution",
       subtitle="Describing cancer deaths rate on NYC community district level",
       caption="NYC Environment & Health Data Portal") +

  theme(legend.position = "bottom",
        legend.title = element_text(size=10),
        legend.text = element_text(size=8),
        plot.title = element_text(face="bold", size=15),
        plot.subtitle = element_text(size=10, margin=margin(b=10)),
        plot.caption = element_text(size=8, margin=margin(t=15)),
        plot.margin=unit(c(1,1,1,1),"cm"),
        panel.background = element_rect(fill = "#e9e9e9"),
        axis.text = element_text(size=8),
        axis.title = element_text(size=10),
        axis.title.x = element_text(margin=margin(t=15)),
        axis.title.y = element_text(margin=margin(r=15)))
```
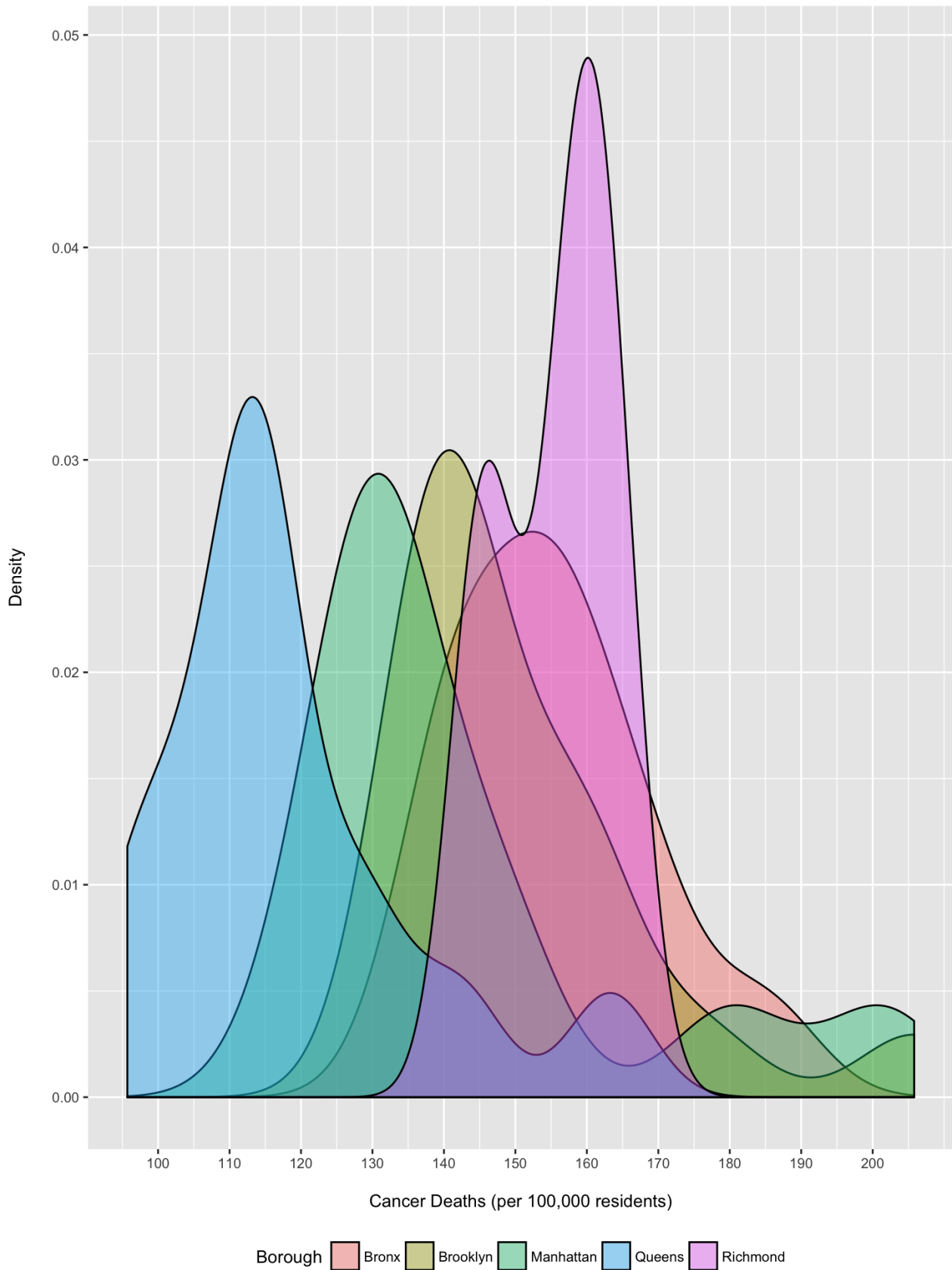
Cancer Deaths Rate Distribution

Describing cancer deaths rate on NYC community district level

NYC Environment & Health Data Portal

Summary The graph is designed to examine the distribution of *Cancer Deaths Rate(2009-2013)*. Note that the death rate for malignant neoplasms (cancers) is adjusted to account for variations in the age structure of the populations. The data are collected on the community district level of NYC, with 59 data points. From the

graph we can see, Bronx, Brooklyn and Manhattan have similar bell-shaped distributions, while Queens is right-skewed, and Richmond has two peaks.