

Problem Set #9 Solutions

Your Name

11/5/2020

Due date: Monday, November 16th

For this problem set we will use data from [New York City](#) that tested children under 6 years old for elevated blood lead levels (BLL). [You can read more about the data on their website].

About the data:

All NYC children are required to be tested for lead poisoning at around age 1 and age 2, and to be screened for risk of lead poisoning, and tested if at risk, up until age 6. These data are an indicator of children younger than 6 years of age tested in NYC in a given year with blood lead levels of 5 mcg/dL or greater. In 2012, CDC established that a blood lead level of 5 mcg/dL is the reference level for exposure to lead in children. This level is used to identify children who have blood lead levels higher than most children's levels. The reference level is determined by measuring the NHANES blood lead distribution in US children ages 1 to 5 years, and is reviewed every 4 years.

Question 1

Recreate the below table with kable.

You will need to calculate the BLL per 1,000, filter for years 2015-2016, and rename the boroughs based on the following coding scheme:

- 1: Bronx
- 2: Brooklyn
- 3: Manhattan
- 4: Queens
- 5: Staten Island

```
bll_nyc_table <- bll_nyc %>%
  mutate(time_period = as.factor(time_period),
         borough_id = as.factor(borough_id)) %>%
  group_by(borough_id, time_period) %>%
  summarize(
    total_tested = sum(total_tested),
    bll_5 = sum(bll_5),
    bll_10 = sum(bll_10),
    bll_15 = sum(bll_15),
    bll_5_1k = round(bll_5/total_tested * 1000, 1),
    bll_10_1k = round(bll_10/total_tested * 1000, 1),
    bll_15_1k = round(bll_15/total_tested * 1000, 1)) %>%
  mutate(borough_id = factor(borough_id,
                             levels = c(1:5),
```

Table 1: BLL Rates per 1,000 tested in New York City, 2015-2016

Borough	Year	BLL >5 µg/dL	BLL >10 µg/dL	BLL >15 µg/dL
Bronx	2015	15.7	2.5	1.0
Bronx	2016	15.0	2.8	1.2
Brooklyn	2015	22.6	3.9	1.3
Brooklyn	2016	22.3	3.6	1.2
Manhattan	2015	10.6	1.6	0.5
Manhattan	2016	8.1	1.3	0.6
Queens	2015	15.4	2.7	1.0
Queens	2016	14.3	2.3	0.9
Staten Island	2015	12.0	2.0	0.7
Staten Island	2016	14.8	2.7	0.8

```

labels = c("Bronx", "Brooklyn",
            "Manhattan", "Queens", "Staten Island"),
ordered=TRUE)) %>%
filter(time_period %in% c("2015", "2016")) %>%
select(borough_id, time_period, bll_5_1k, bll_10_1k, bll_15_1k)

```

'summarise()' regrouping output by 'borough_id' (override with '.groups' argument)

```

kable(bll_nyc_table,
      booktabs=T,
      col.names=c("Borough", "Year", "BLL >5 µg/dL", "BLL >10 µg/dL", "BLL >15 µg/dL"),
      align='lccc',
      caption="BLL Rates per 1,000 tested in New York City, 2015-2016",
      format.args=list(big.mark=","))

```

Question 2

Replicate the following bar chart. Since we want the graph to have an ascending order, we will need to factor borough_id. <https://colorswall.com/palette/454/>:

- #ff6600: orange
- #003884: blue

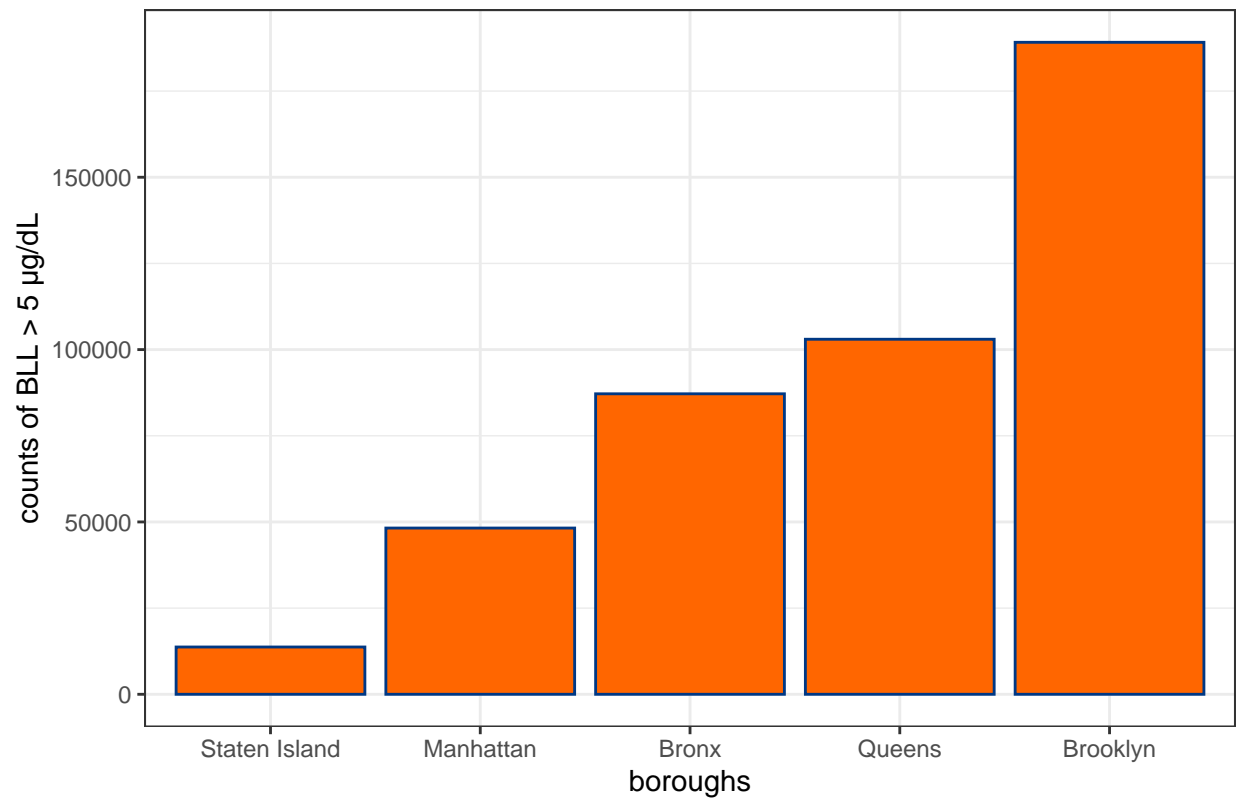
```
bll_nyc_1 <- bll_nyc %>%
  group_by(borough_id) %>%
  summarize(bll_all = sum(bll_5, bll_10, bll_15))

## 'summarise()' ungrouping output (override with '.groups' argument)

bll_nyc_1$borough_id <- factor(bll_nyc_1$borough_id,
                              levels = c(5,3,1,4,2), # "2", "3", "1", "4", "5"
                              labels = c("Staten Island", "Manhattan",
                                           "Bronx", "Queens", "Brooklyn"),
                              ordered=TRUE)

ggplot(bll_nyc_1, aes(x = borough_id, y = bll_all)) +
  geom_col(fill = "#ff6600", color = "#003884") +
  labs(x = "boroughs",
       y = "counts of BLL > 5 µg/dL",
       title = "New York City: Elevated Blood Lead Levels 2005-2016 by Borough") +
  #scale_x_discrete(labels=c("Bronx", "Brooklyn", "Manhattan",
                           # "Queens", "Staten Island")) +
  theme_bw()
```

New York City: Elevated Blood Lead Levels 2005–2016 by Borough



```
#ggsave("data/question_2_bar.png")
```

Question 3

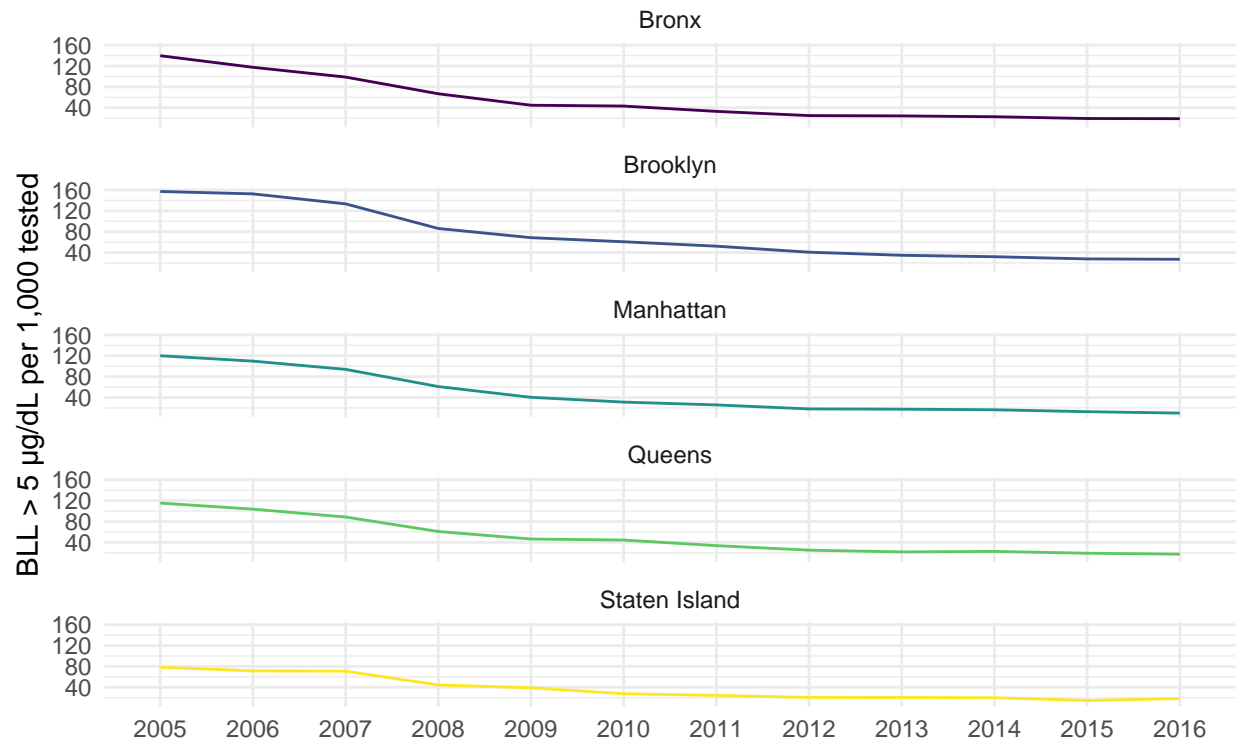
Recreate the following line graph. You will need to combine the three measures of BLL to calculate a rate per borough per year. Your data frame, at it's smallest, will have 60 observations (12 years x 5 boroughs).

```
bll_nyc_viz <- bll_nyc %>%
  mutate(time_period = as.factor(time_period),
         borough_id = as.factor(borough_id)) %>%
  group_by(borough_id, time_period) %>%
  summarize(bll = sum(bll_5, bll_10, bll_15), # get total bll
            total_tested = sum(total_tested), # sum of total tested per row
            bll_5plus_1k = round(bll/total_tested * 1000, 1)) %>%
  mutate(borough_id = factor(borough_id,
                             levels = c(1:5),
                             labels = c("Bronx", "Brooklyn", "Manhattan",
                                         "Queens", "Staten Island"),
                             ordered=TRUE))

## 'summarise()' regrouping output by 'borough_id' (override with '.groups' argument)

ggplot(bll_nyc_viz, aes(x = time_period, y = bll_5plus_1k, group = borough_id)) +
  facet_wrap(~ borough_id, ncol=1) +
  geom_line(aes(color = borough_id)) +
  theme_minimal() +
  labs(y = "BLL > 5 µg/dL per 1,000 tested",
       title = "New York City: Elevated Blood Levels > 5 µg/dL",
       subtitle = "2005-2016") +
  theme(legend.position="none", axis.title.x = element_blank())
```

New York City: Elevated Blood Levels > 5 $\mu\text{g/dL}$ 2005–2016



```
#ggsave("data/question_3_line.png")
```