Lab 12 - Statistics, Coordinates, Facets, and Themes

Your Name Here

November 14, 2017

Complete the following exercises below. Knit together the PDF document and commit both the Lab 12 RMDfile and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Choose one or more graphics you created for Lab 11 and either experiment with the underlying statistical layer if it already has one (i.e. if you made a histogram experiement with different bin widths) or add a separate statistical layer to your plot (i.e. a smooothing curve). Choose something you think will offer meaningful insight and describe why you made the choice you did. What additional information does this provide viewers of your graphic?

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
setwd("/Users/marakage/Desktop/Honors/STAT/Labs")
AsianAlone <- read.csv("DEC_10_SF2_PCT1_with_ann.csv")
Totalpopulation <- read_csv("../Datasets/TotalPopulation/ACS_10_5YR_B01003_with_ann.csv",skip=0)
```

```
## Parsed with column specification:
## cols(
##   GEO.id = col_character(),
##   GEO.id2 = col_character(),
##   `GEO.display-label` = col_character(),
##   HD01_VD01 = col_character(),
##   HD02_VD01 = col_character()
## )
```

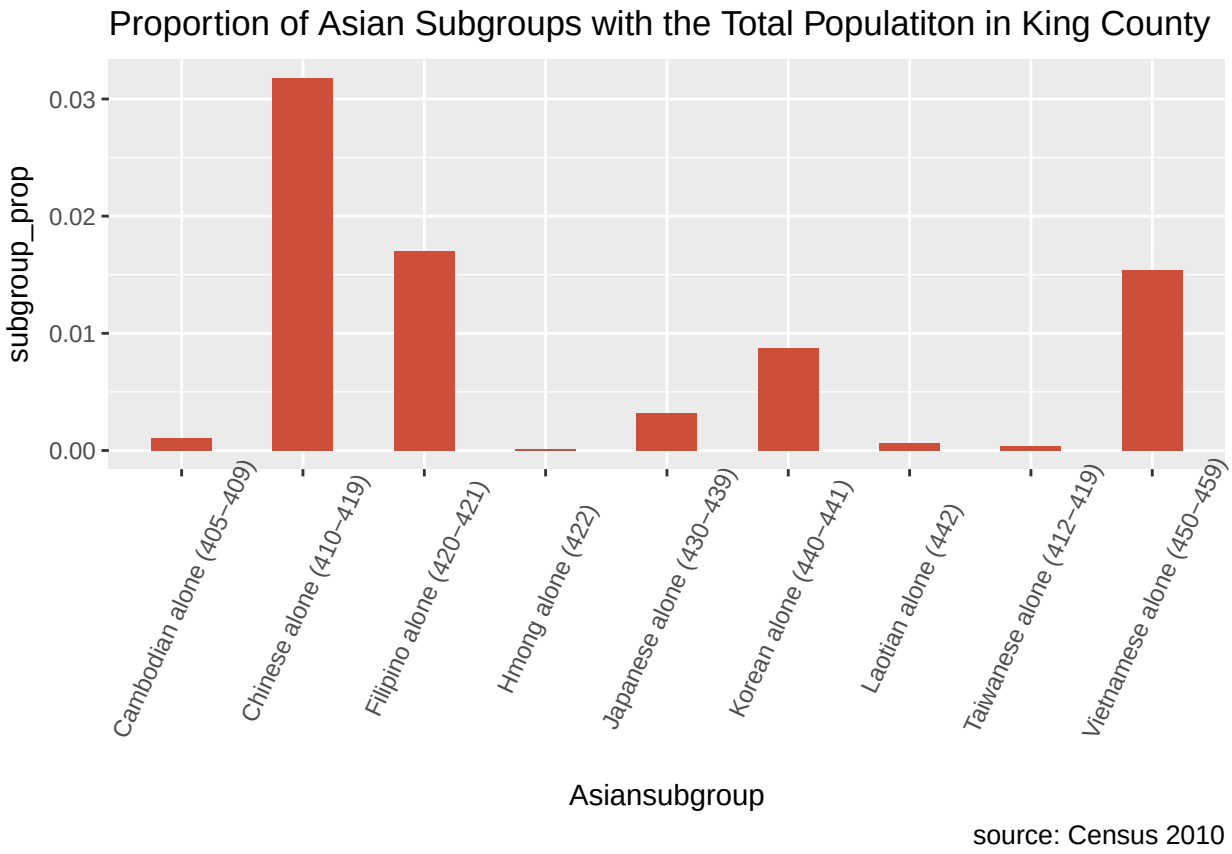
```
data_asian_alone <- read_csv("./asian/data_asian_alone.csv")
```

```
## Parsed with column specification:
## cols(
##   POPGROUP.id = col_character(),
##   POPGROUP.display.label = col_character(),
##   GEO.id = col_character(),
##   GEO.id2 = col_double(),
##   GEO.display.label = col_character(),
##   D001 = col_integer(),
##   Count = col_integer()
## )
```

```r
#AsianAlone <- read.csv("DEC_10_SF2_PCT1_with_ann.csv")
data_subset_total <- Totalpopulation %>%
  filter(GEO.id != "Id") %>%
  mutate(TotalCount = as.numeric(as.character(HD01_VD01)))

data_asian_alone <- data_asian_alone %>%
  mutate(GEO.id2 = as.character(GEO.id2))

mergedata <- data_subset_total %>%
  left_join(data_asian_alone, by="GEO.id2") %>%
  select(-GEO.id.y, -ends_with("label.y")) %>%
  mutate(totalprop = Count/ TotalCount)
submergedata <- mergedata %>%
  filter(POPGROUP.id != "012" & POPGROUP.id != "031")

#total number of tracts that doesn't have Asian Alone or in combination
sum(is.na(mergedata$Count))
```

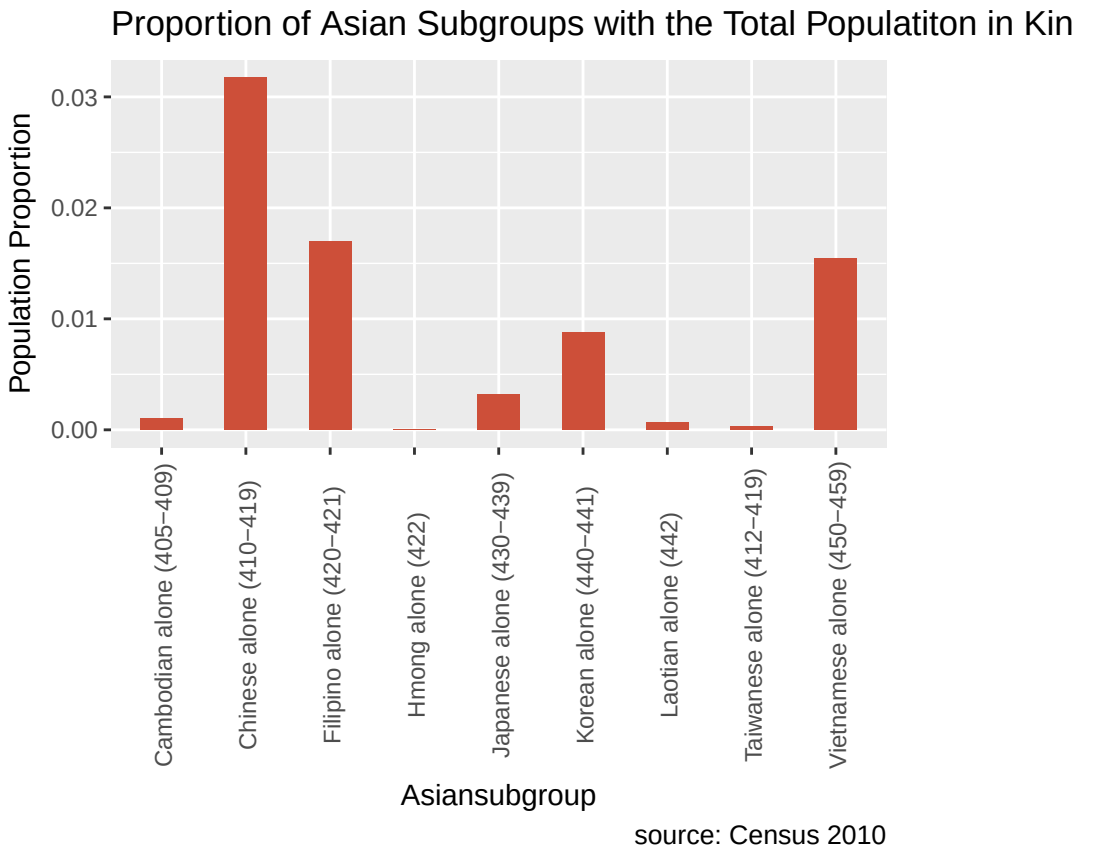
```
## [1] 117
```

```r
barpop <- submergedata %>%
  mutate(Asiansubgroup = `POPGROUP.display.label`) %>%
  group_by(POPGROUP.id, Asiansubgroup) %>% #join all the tract population by same subgroup
  summarise(subpoptotal = sum(Count)) %>%
  mutate(subgroup_prop = round(subpoptotal/sum(data_subset_total$TotalCount), 5))#new column with the p


ggplot(barpop, aes(x=Asiansubgroup, y=subgroup_prop)) +
  geom_bar(stat="identity", width=.5, fill="tomato3") +
  labs(title="Proportion of Asian Subgroups with the Total Populatiton in King County",
       caption="source: Census 2010") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))
```



2. With the same or a different plot created in Lab 11, experiment with zooming in on specific areas of your graphic and changing the aspect ratio. Are their any benefits/drawbacks with either or both of these approaches for the visualizations you've created? What are they?

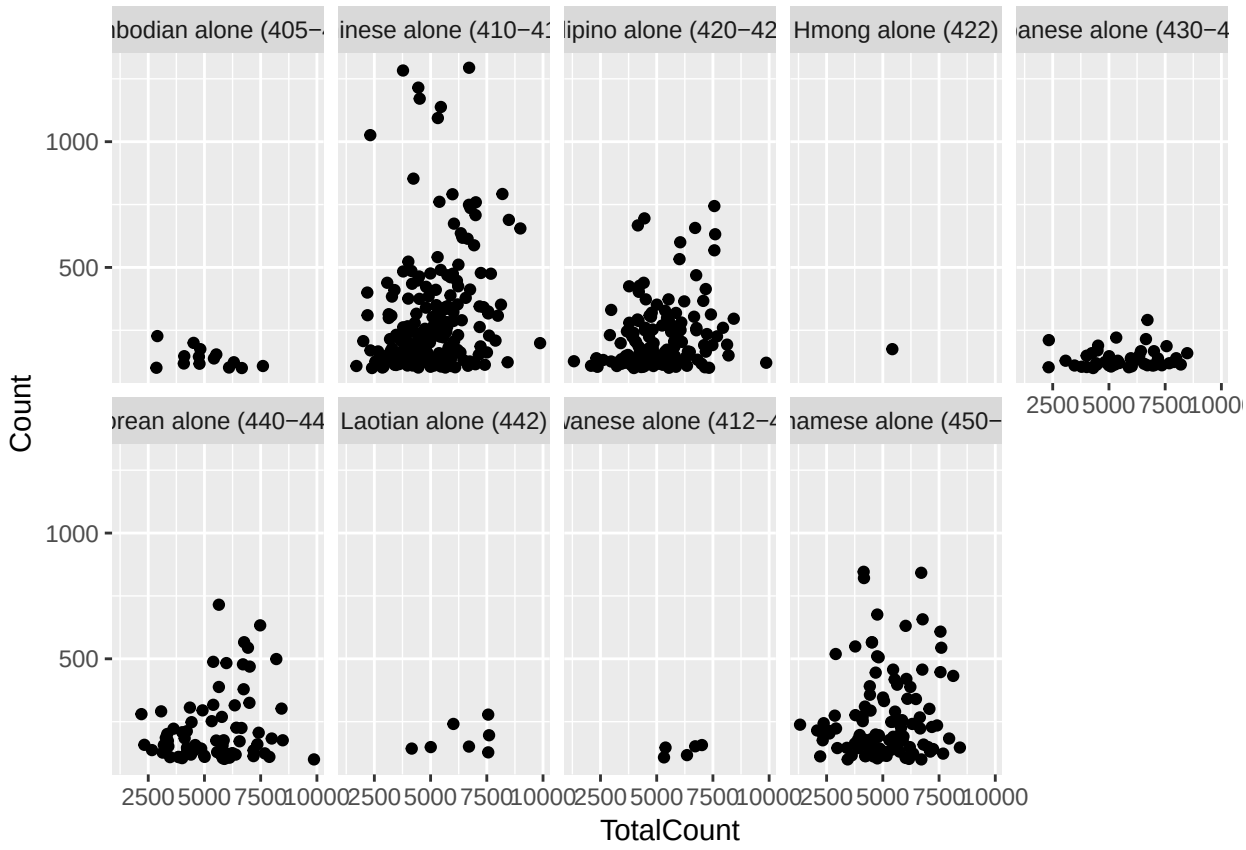
```
ggplot(barpop, aes(x=Asiansubgroup, y=subgroup_prop)) +
  geom_bar(stat="identity", width=.5, fill="tomato3") +
  labs(title="Proportion of Asian Subgroups with the Total Populatiton in King County",
       caption="source: Census 2010") +
  theme(axis.text.x = element_text(angle=90, vjust=0.6), aspect.ratio = 0.5) +
  ylab("Population Proportion")
```



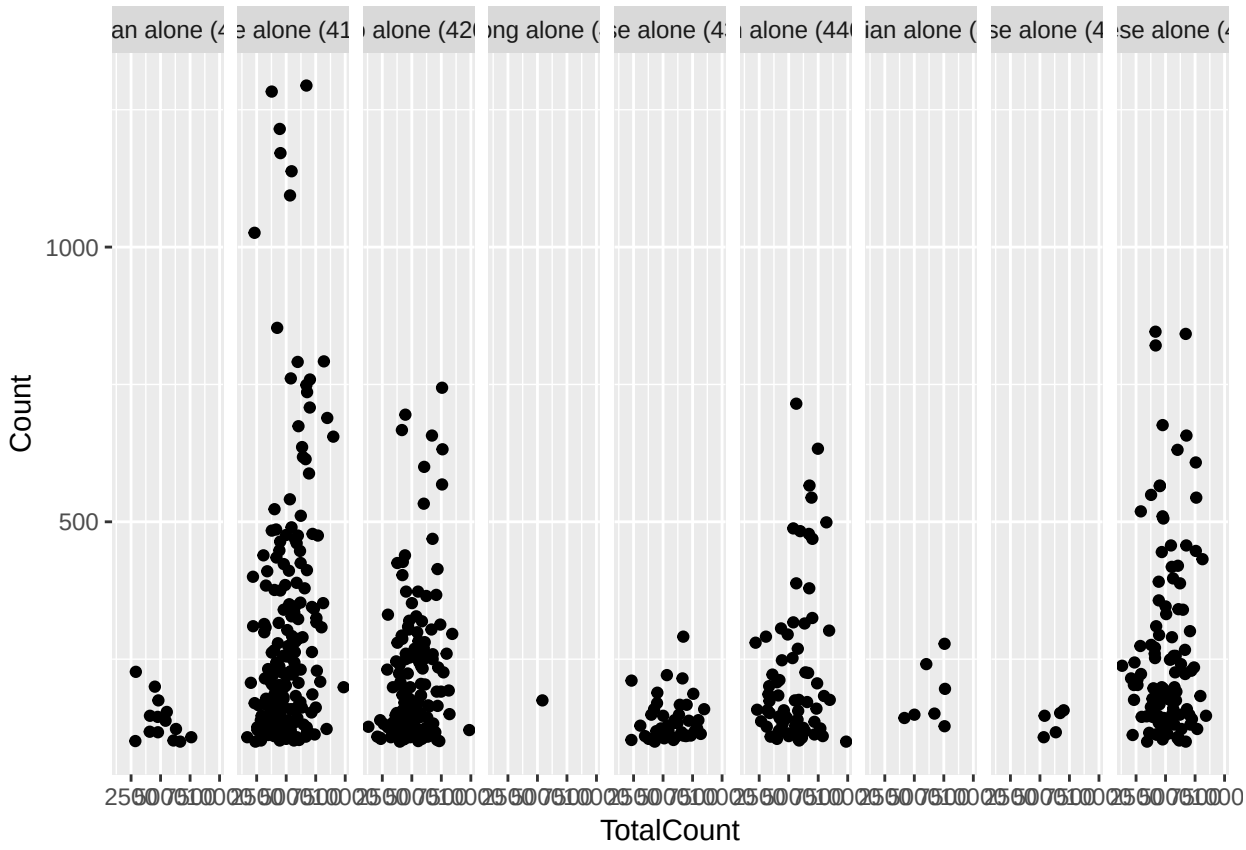

```
#yes, the aspect ration now emphasized the x-axis and helped visualize the subgroups bettter.
```

3. Try facetting a plot you have made by another categorical variable in your data (this can even be as simple as Male/Female). What is the difference between `facet_wrap()` and `facet_grid()`? How might facetting be useful in data visualization?

```
ggplot(data = submergedata) +
  geom_point(mapping = aes(x = TotalCount, y = Count)) +
  facet_wrap(~ `POPGROUP.display.label`, nrow = 2)
```




```
ggplot(data = submergedata) +
  geom_point(mapping = aes(x = TotalCount, y = Count)) +
  facet_grid(~ `POPGROUP.display.label`)
```

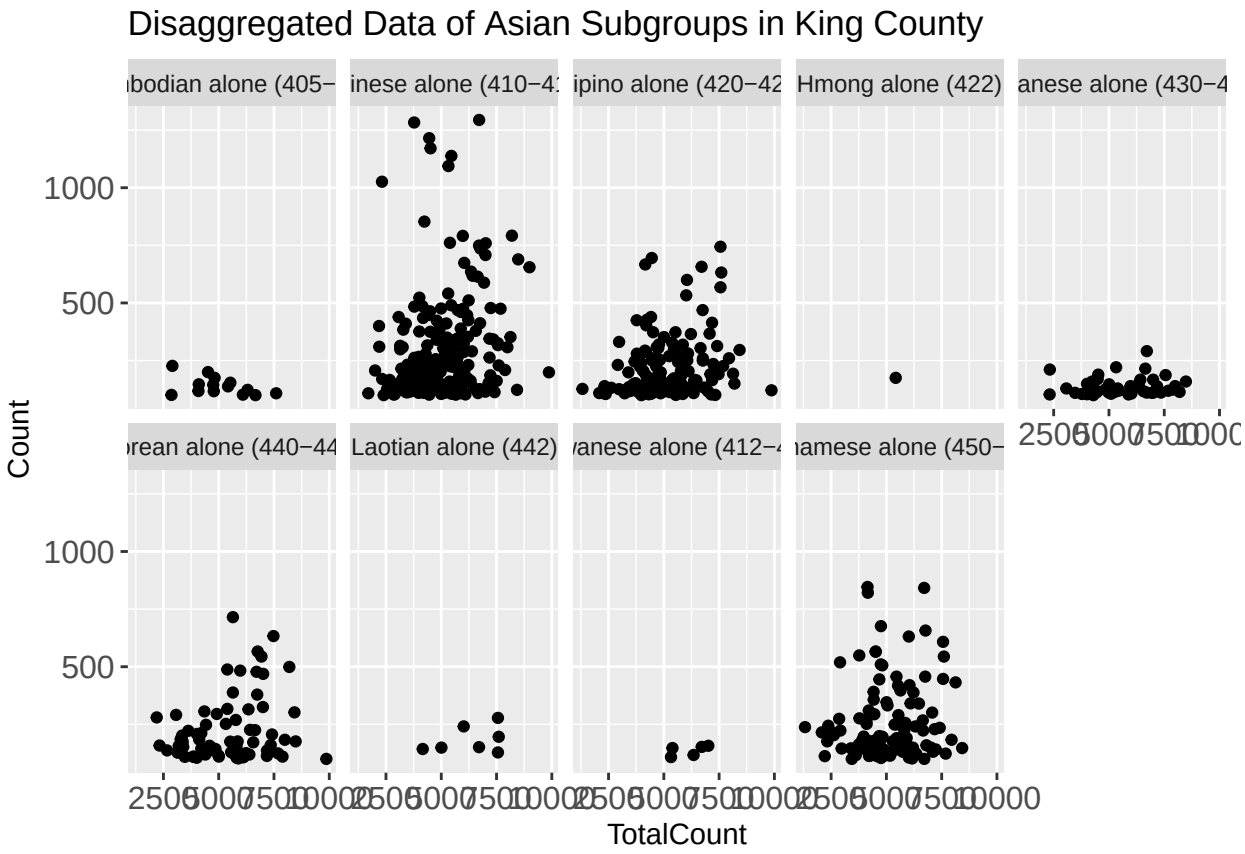



```
#We can use small multipoles to compare and contrast diffrent groups. Faced grid didn't help in this ca
```

4. Use the `theme()` layer to change the appearance of a plot of your choice including the

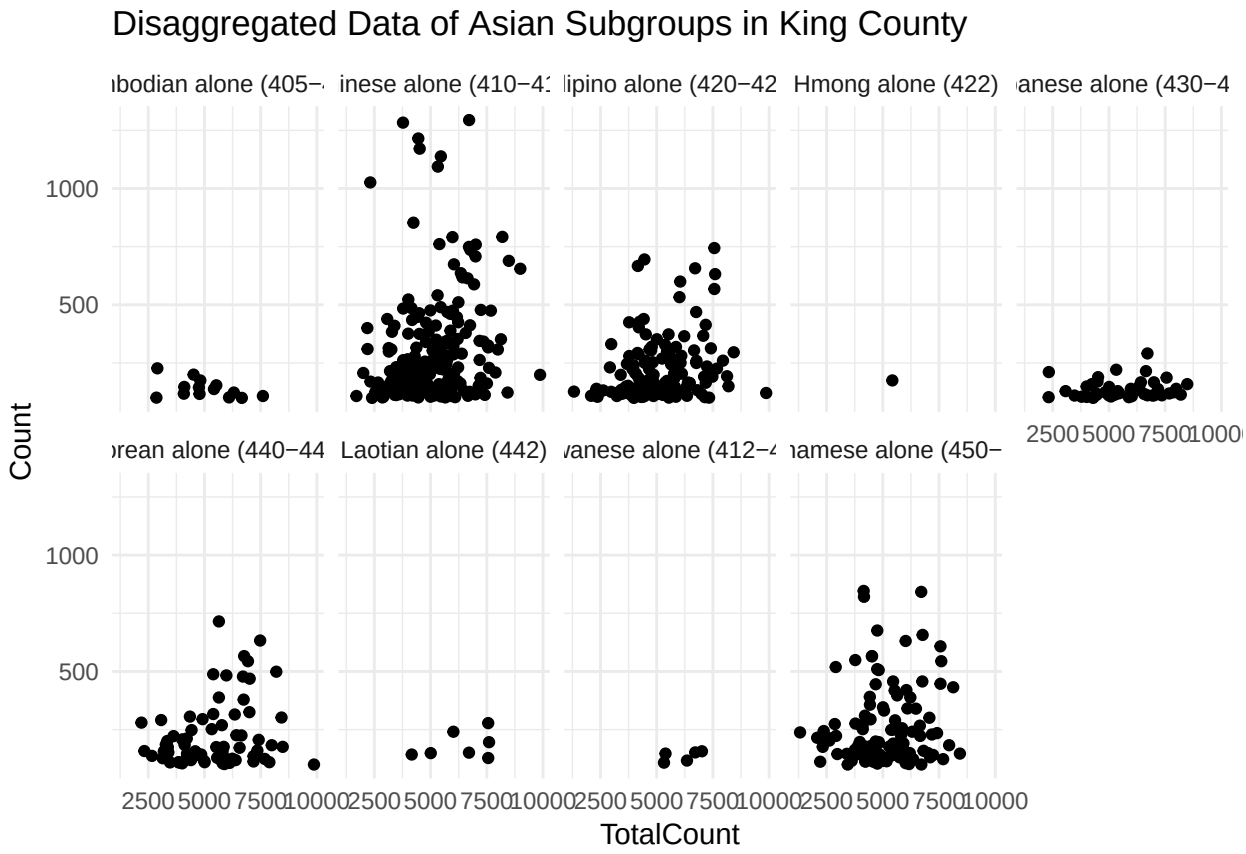
- plot, axes, and legend titles
- axes tick marks
- text size
- legend position

```
ggplot(data = submergedata) +
  geom_point(mapping = aes(x = TotalCount, y = Count)) +
  facet_wrap(~ `POPGROUP.display.label`, nrow = 2) +
  theme(axis.text = element_text(size = 12)) +
  labs(title = "Disaggregated Data of Asian Subgroups in King County")
```

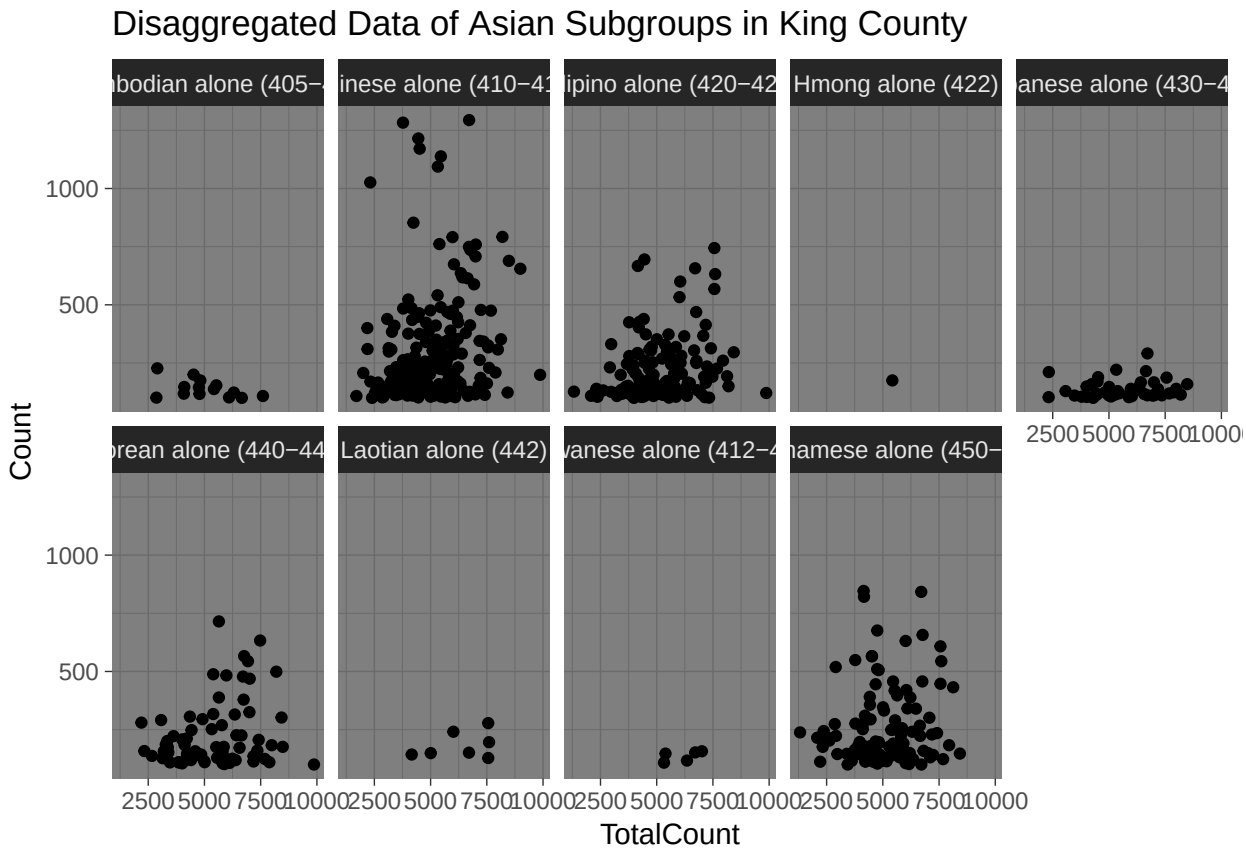


5. Create three versions of a graphic of your choice using different built-in themes or a theme created from `ggthemes`. Which ones do you think are best for presenting in an academic journal? A poster session? What are the qualities of the themes that you choice that you think make them more appropriate for presentation?

```
ggplot(data = submergedata) +
  geom_point(mapping = aes(x = TotalCount, y = Count)) +
  facet_wrap(~ `POPGROUP.display.label`, nrow = 2) +
  theme_minimal() +
  labs(title = "Disaggregated Data of Asian Subgroups in King County")
```




```
ggplot(data = submergedata) +
  geom_point(mapping = aes(x = TotalCount, y = Count)) +
  facet_wrap(~ `POPGROUP.display.label`, nrow = 2) +
  theme_dark() +
  labs(title = "Disaggregated Data of Asian Subgroups in King County")
```




```
ggplot(data = submergedata) +
  geom_point(mapping = aes(x = TotalCount, y = Count)) +
  facet_wrap(~ `POPGROUP.display.label`, nrow = 2) +
  ggthemes::theme_map() +
  labs(title = "Disaggregated Data of Asian Subgroups in King County")
```

Disaggregated Data of Asian Subgroups in King County