

FinalReport

Karan Dassi

5/28/2019

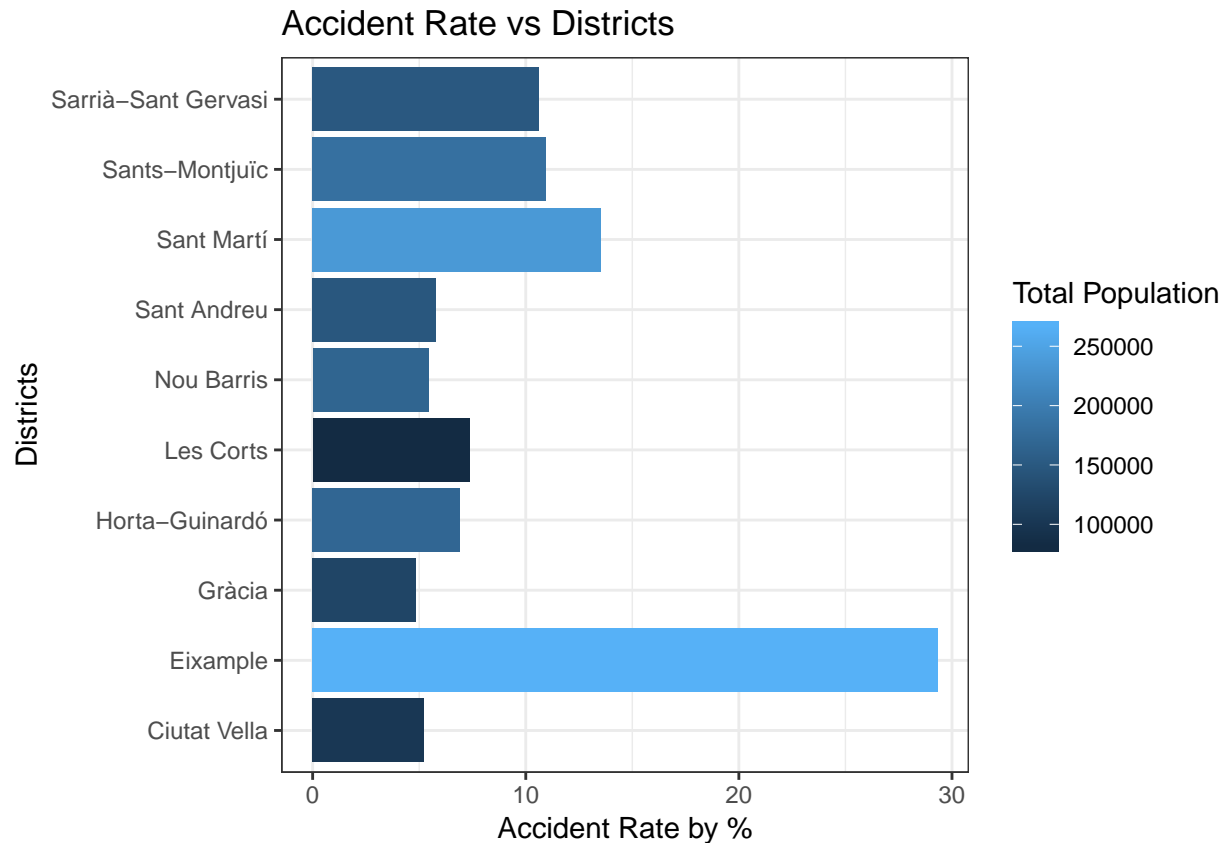
Accident Rate vs Population by District

```
accidents%>%
  group_by(DISTRICT.NAME)%>%
  summarise(total_vic = sum(VICTIMS, na.rm = T))%>%
  mutate(vic_rate = round(100 * total_vic / sum(total_vic),2))%>%
  arrange(desc(total_vic)) -> accidents_vic

population%>%
  group_by(DISTRICT.NAME)%>%
  summarise(total_pop = round(sum(NUMBER_P, na.rm = T))) -> dis_pop

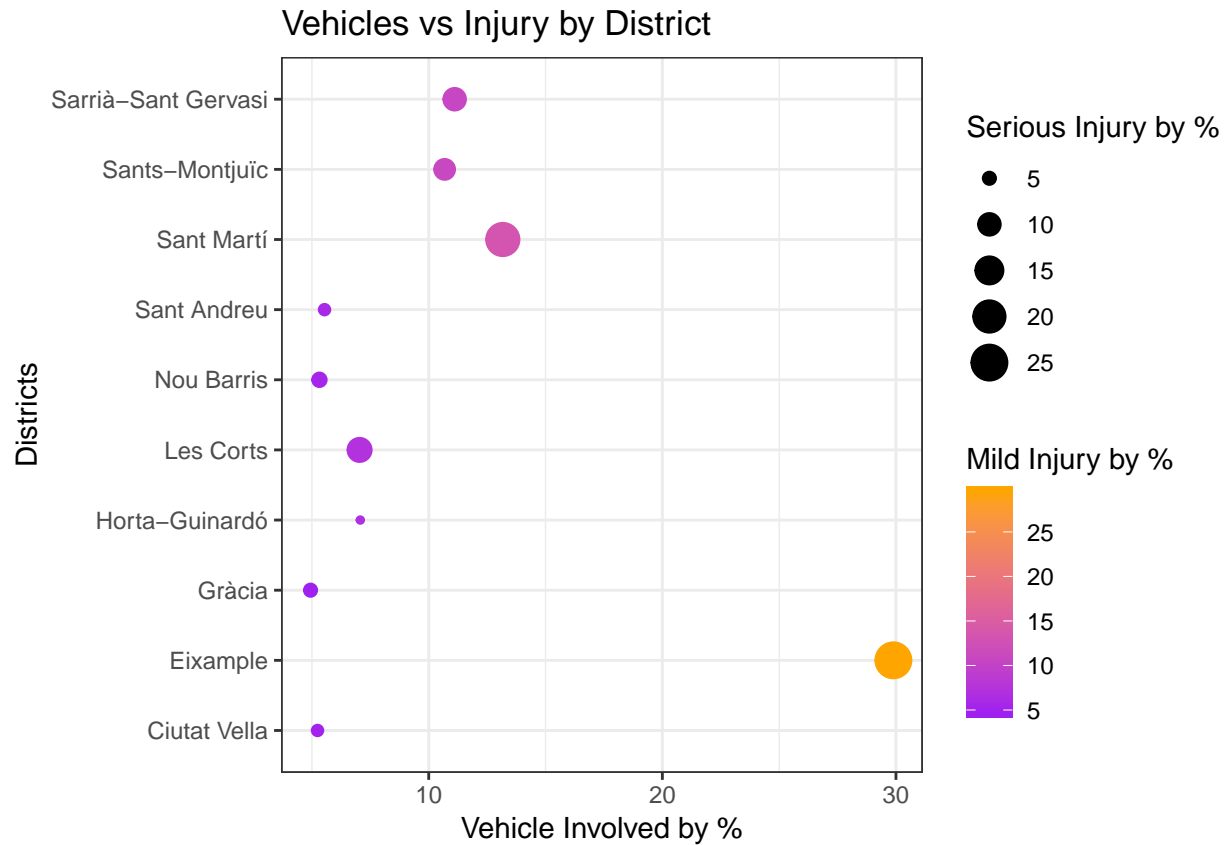
left_join(accidents_vic, dis_pop, by = "DISTRICT.NAME")%>%
  arrange(desc(vic_rate)) -> dis_vic_pop

dis_vic_pop%>%
  ggplot(aes(x = DISTRICT.NAME, y = vic_rate, fill = total_pop)) +
  geom_col() +
  theme_bw() +
  coord_flip() +
  ggtitle("Accident Rate vs Districts") +
  xlab("Districts") +
  ylab("Accident Rate by %") +
  labs(fill = "Total Population")
```



Vehicles Involved vs Injury Level by Districts

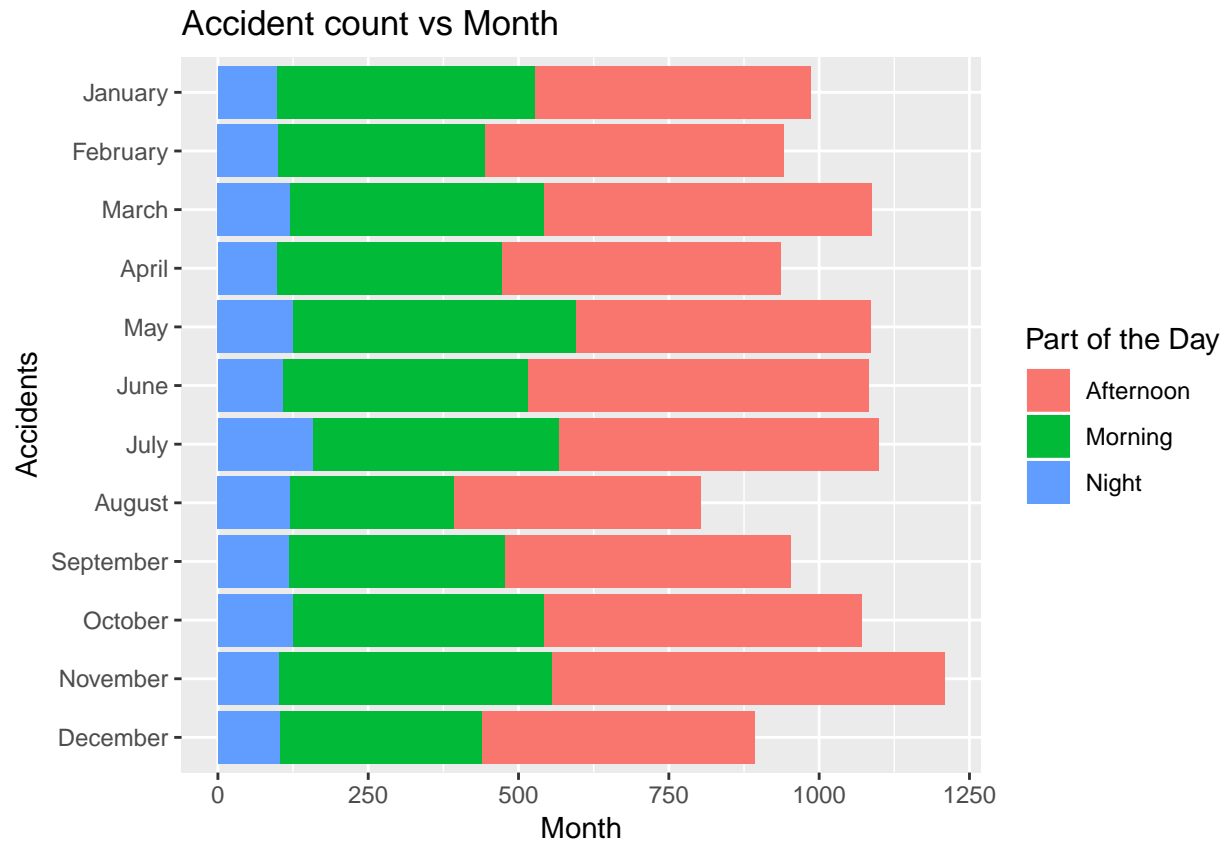
```
accidents%>%
  group_by(DISTRICT.NAME)%>%
  summarise(sum_veh = sum(VEHICLES.INVOLVED),
            sum_mid = sum(MILD.INJURIES),
            sum_ser = sum(SERIOUS.INJURIES))%>%
  mutate(mid_rate = round(100 * sum_mid / sum(sum_mid), 2),
         ser_rate = round(100 * sum_ser / sum(sum_ser), 2),
         veh_rate = round(100 * sum_veh / sum(sum_veh), 2))%>%
  ggplot(aes(x = DISTRICT.NAME, y = veh_rate, color = mid_rate, size = ser_rate)) +
  geom_point() +
  scale_colour_gradient(low = "Purple", high = "orange") +
  coord_flip() +
  theme_bw() +
  ggtitle("Vehicles vs Injury by District") +
  xlab("Districts") +
  ylab("Vehicle Involved by %") +
  labs(color = "Mild Injury by %", size = "Serious Injury by %")
```



Accident Count based on Month & Time of Day

```
monthvec <- c("December", "November", "October", "September", "August", "July", "June", "May", "April",
              "March", "February", "January")

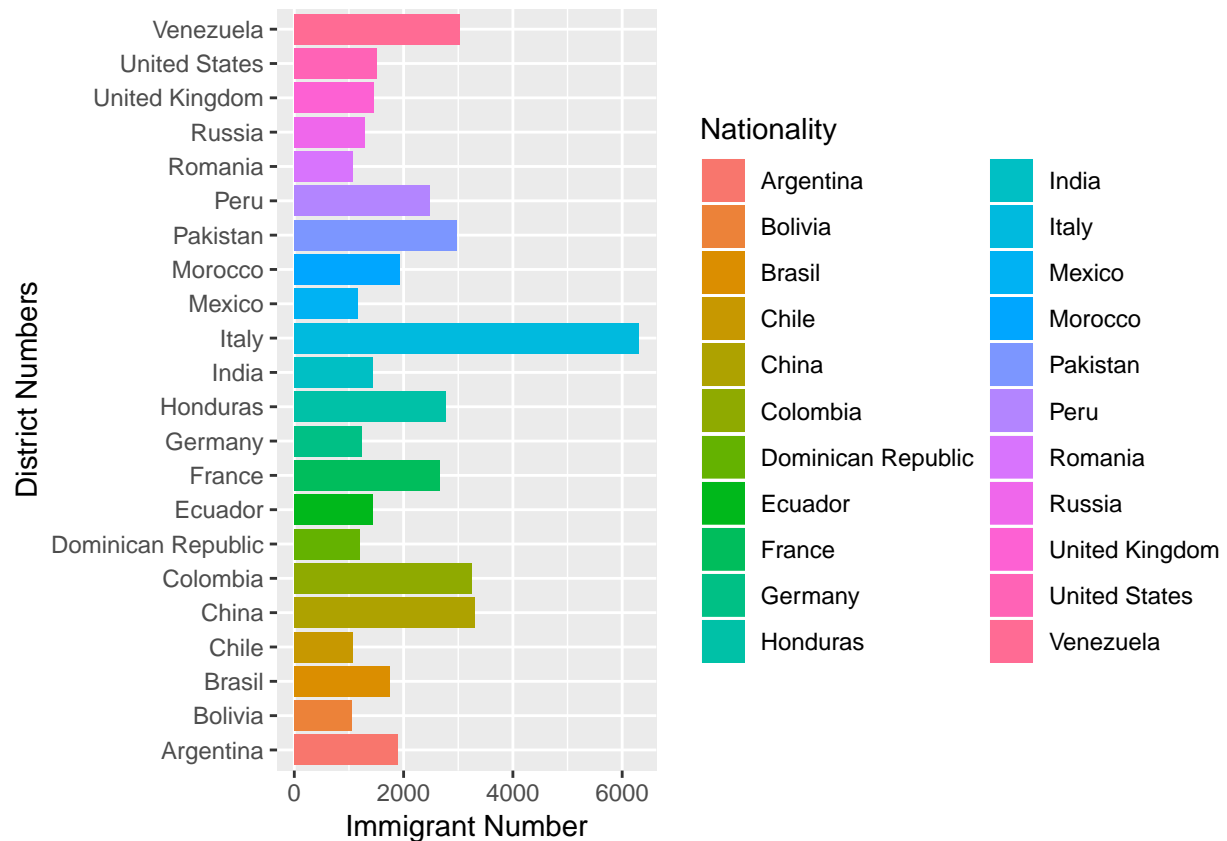
accidents%>%
  mutate(month_fc = parse_factor(MONTH, levels = monthvec))%>%
  group_by(month_fc, PART.OF.THE.DAY)%>%
  summarise(total_vic = sum(VICTIMS))%>%
  ggplot(aes(x = month_fc, y = total_vic, fill = PART.OF.THE.DAY)) +
  geom_col() +
  coord_flip() +
  ggtitle("Accident count vs Month") +
  xlab("Accidents") +
  ylab("Month") +
  labs(fill = "Part of the Day")
```



Immigration vs Nationality

```
immigration_nationality %>%
  filter(Year == 2017) %>%
  select(Year, District.Name, Nationality, Number) %>%
  group_by(Nationality) %>%
  summarize(TotalNumber = sum(Number)) %>%
  arrange(desc(TotalNumber)) %>%
  filter(TotalNumber > 1000 & Nationality != "Spain") ->tidy_immigration_nationality

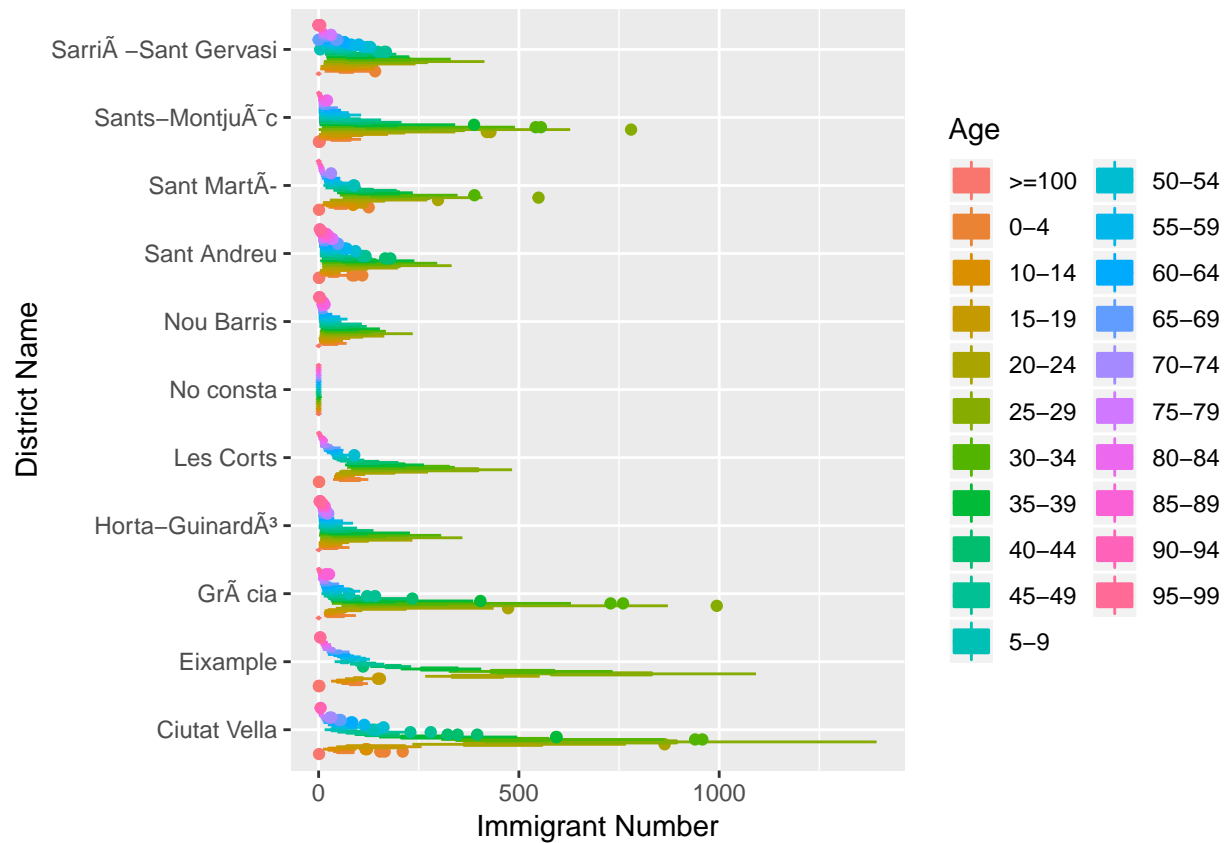
ggplot(tidy_immigration_nationality, aes(x =Nationality, y = TotalNumber, fill = Nationality)) +
  geom_bar(stat = "identity") +
  xlab("District Numbers") +
  ylab("Immigrant Number") +
  coord_flip()
```



```
immigration_age %>%
  select(Year, District.Name, Age, Immigrants) %>%
  group_by(Age) %>%
  summarize(TotalNumber = sum(Immigrants)) %>%
  arrange(desc(TotalNumber)) %>%
  filter(TotalNumber > 1000) ->tidy_immigration_age
```

Immigrant Age Groups Analysis:

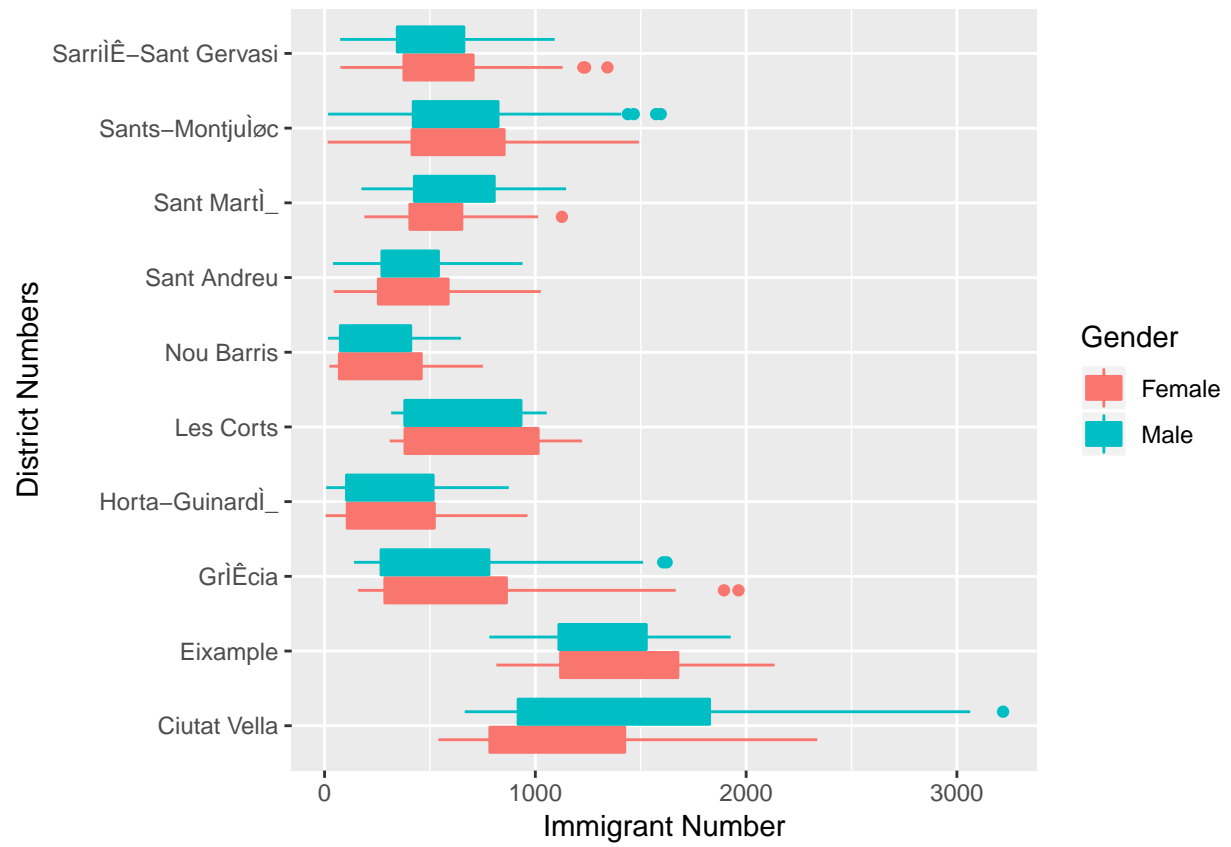
```
ggplot(immigration_age, aes(x =as.character(District.Name), y = Immigrants, color = Age, fill = Age)) +
  geom_boxplot() +
  xlab("District Name") +
  ylab("Immigrant Number") +
  coord_flip()
```



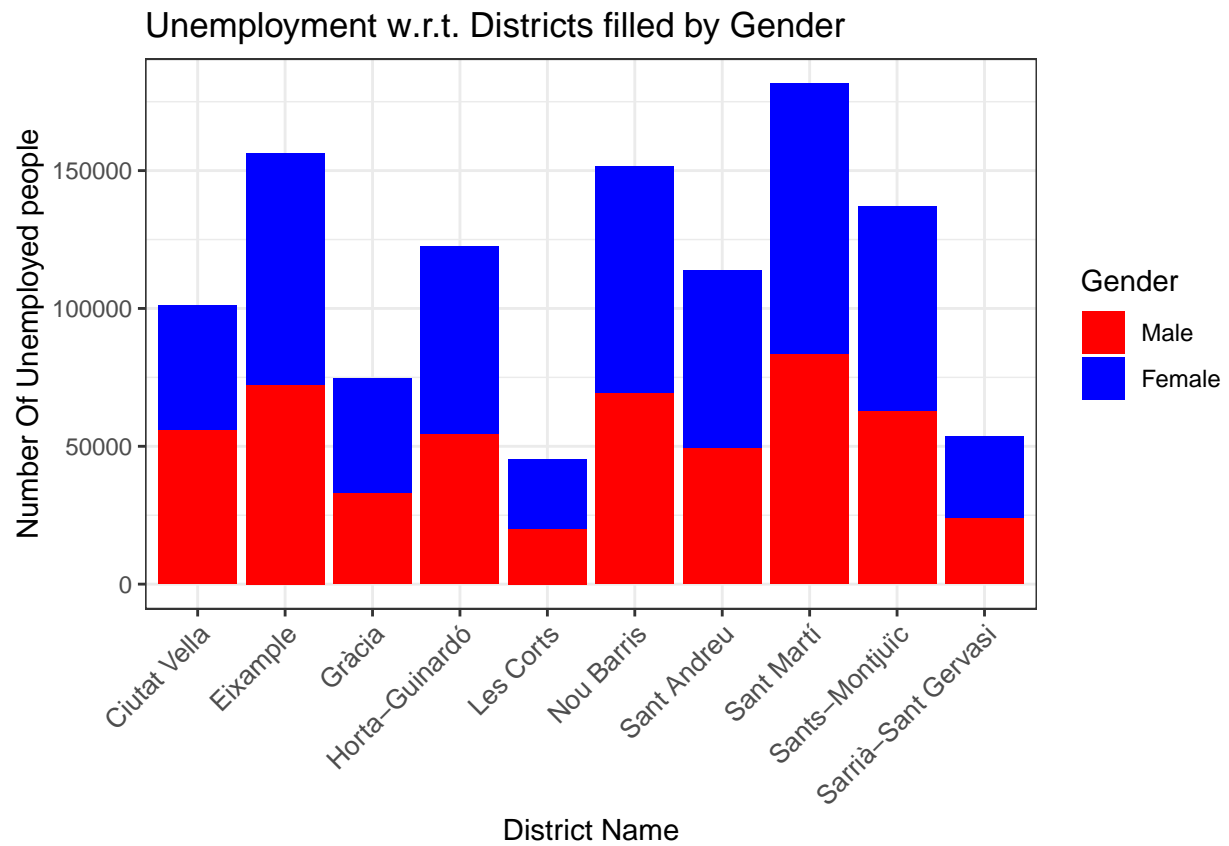
```
immigration_sex %>%
  select(Year, District.Name, Gender, Immigrants) %>%
  group_by(Gender) %>%
  summarize(TotalNumber = sum(Immigrants)) %>%
  arrange(desc(TotalNumber)) %>%
  filter(TotalNumber > 1000) ->tidy_immigration_sex
```

Immigrant Gender Analysis:

```
ggplot(immigration_sex, aes(x = as.character(District.Name), y = Immigrants, color = Gender, fill = Gender)) +
  geom_boxplot() +
  xlab("District Numbers") +
  ylab("Immigrant Number") +
  coord_flip()
```



Gender Wise Unemployment



Unemployment vs Accidents w.r.t. Districts

