Data Sci: Group 10 Data Brainstorm

Daniel Forcade

3/11/2021

Names of Variables in Data Set:

```
allegations <- read.csv("allegations_202007271729.csv")

names(allegations)
```

```
##  [1] "unique_mos_id"           "first_name"
##  [3] "last_name"               "command_now"
##  [5] "shield_no"               "complaint_id"
##  [7] "month_received"          "year_received"
##  [9] "month_closed"            "year_closed"
## [11] "command_at_incident"     "rank_abbrev_incident"
## [13] "rank_abbrev_now"         "rank_now"
## [15] "rank_incident"           "mos_ethnicity"
## [17] "mos_gender"              "mos_age_incident"
## [19] "complainant_ethnicity"   "complainant_gender"
## [21] "complainant_age_incident" "fado_type"
## [23] "allegation"              "precinct"
## [25] "contact_reason"          "outcome_description"
## [27] "board_disposition"
```

1.

Ran a histogram of complaint type, separated by year. Noticed that years 1985 ~ 2000 have a limited amount of data. May need to research as to why.

```
### Complaint Type Histogram
ggplot(data = allegations,
       mapping = aes(x = fado_type, fill = fado_type, alpha = .8))+
  geom_histogram(stat = "count")+
  facet_wrap(~ year_received)+
  labs(title = "Complaint Type Histogram through Year", x = "Complaint Type", y = "Count", caption = "F
  theme_dark()+
  theme(axis.text.x = element_text(angle = 90, vjust = -.1))+
  scale_fill_tron(name = "Complaint Type")+
  guides(alpha = FALSE)
```
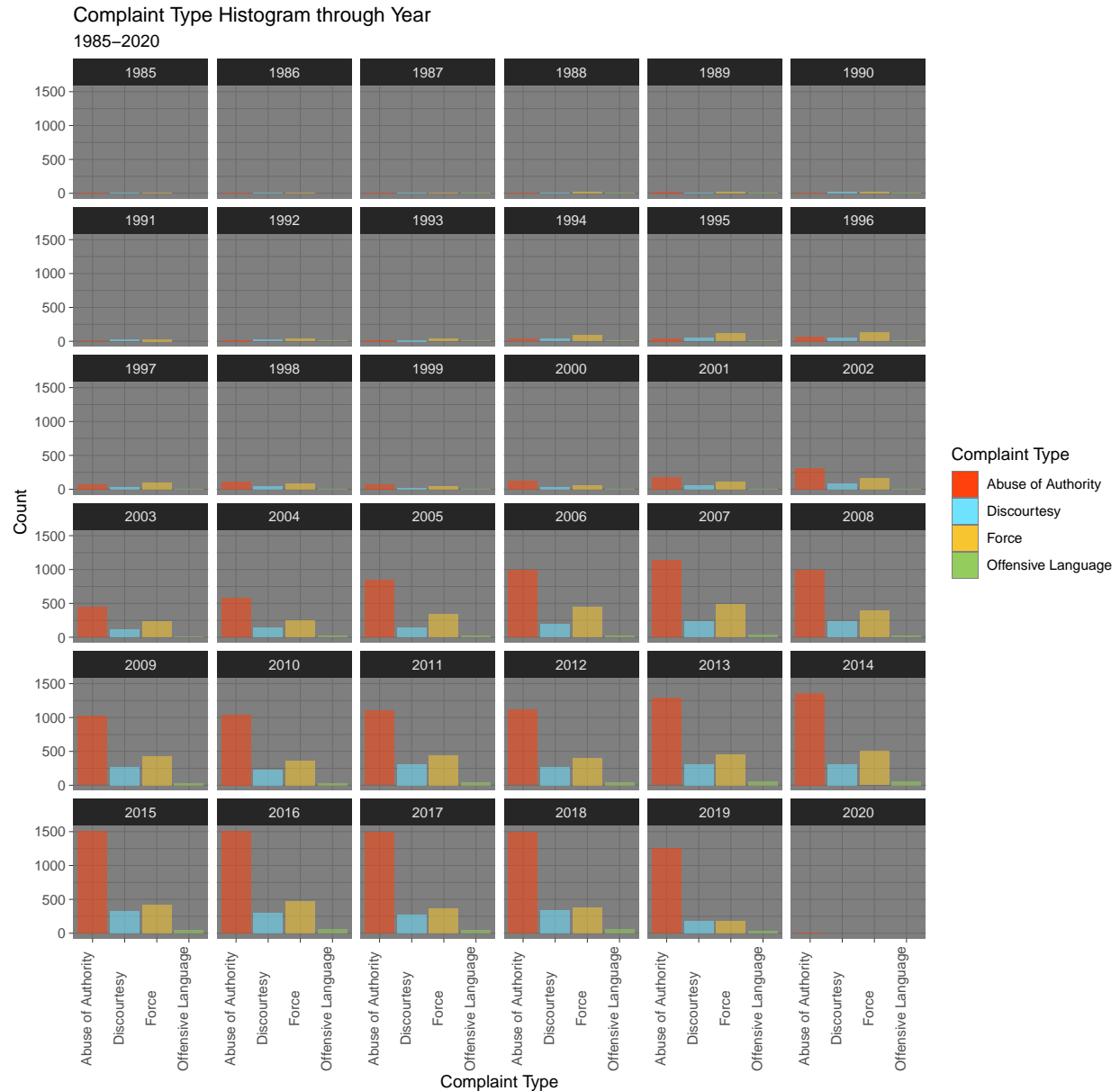
Complaint Type Histogram through Year
1985–2020

Figure 1

2.

Figure 2 is a density bar graph of complaint ethnicity, separated by complaint type. I wanted to see if there were any trends in complaints depending on ethnicity of the victim.

I noticed in this graph that there are a lot of _____ or 'blank', 'unfilled' categories in this data set. These are in addition to several 'NA' or 'Unknown' categories. We will likely have to decide on a path on how to rewrite, categorize, or otherwise parse the missing data.

```
## complaint ethnicity
ggplot(data = allegations,
       mapping = aes(x = complainant_ethnicity, fill = complainant_ethnicity,
                     alpha = .9))+
  geom_bar(color = "black")+
```

```
facet_wrap(~fado_type)+
labs(title = "Complaintant Ethnicity Bar Chart", subtitle = "Seperated by Complaint Type.  Note: First
theme_dark()+
theme(axis.text.x = element_text(angle = 90))+
scale_fill_discrete(name = "Complaint Ethnicity")+
guides(alpha = FALSE)
```

Complaintant Ethnicity Bar Chart
Seperated by Complaint Type. Note: First Value is Blank
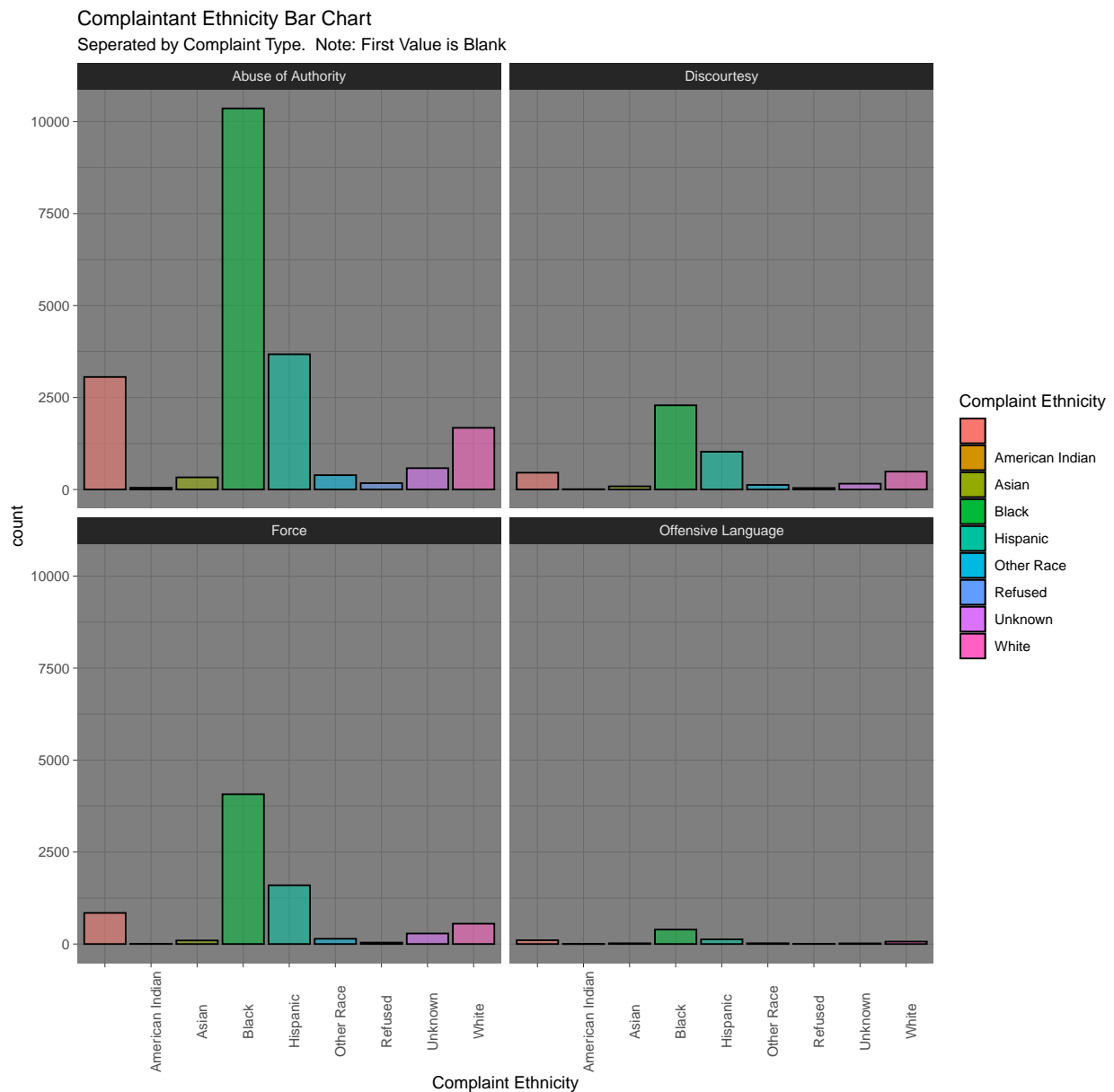
Figure 2

3.

Figure 3 is a density bar graph of `board_disposition`, which I believe to be the outcome of the complaint
- separated by `mos_ethnicity`, which I believe to be officer ethnicity.
A quick research into the data lead me to believe that 'mos' in this data set stands for **Member of Service**,

or police officer.

Here I was checking for patterns if there were any obvious outcome bias depending on officer ethnicity.

```
## Officer Race by Board Disposition
ggplot(data = allegations,
       mapping = aes(x = board_disposition, fill = board_disposition))+
  geom_bar(color = "black")+
  facet_wrap(~mos_ethnicity)+
  theme_cleveland()+
  theme(axis.text.x=element_blank(),
        axis.ticks.x=element_blank())+
  labs(title = "Officer Complaint Outcome Bar Chart", subtitle = "Seperated by Officer Ethnicity", capti
  scale_fill_futurama(name = "Complaint Outcome")
```
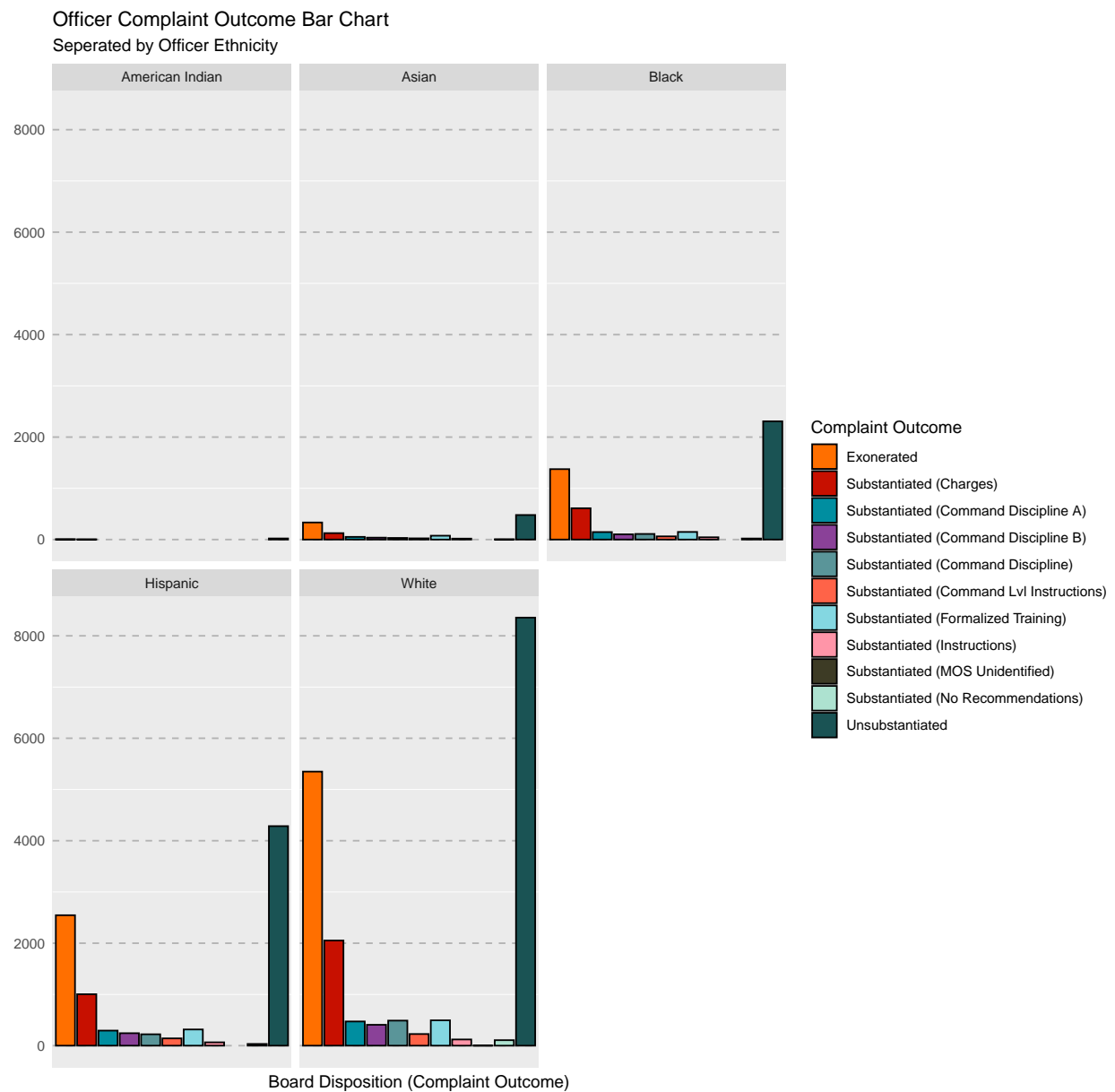

Figure 3

4

4.

Here I did a quick scan on compliant type by `complainant_gender`. As the graph-set shows, there are a lot of different and unknown gender variables for complainants. If we want to work with complainant gender, we may want to limit or parse the 4 low-value and/or missing categories (Transman, Not described, Transwoman, Gender non-conforming)

```
## Gender Complaints

ggplot(data = allegations,
       mapping = aes(x = fado_type, fill = fado_type,
                     alpha = .9))+
  geom_bar(color = "black")+
  facet_wrap(~complainant_gender)+
  labs(title = "Complaint Type", subtitle = "Seperated by Gender", caption = "Figure 4", x = "Complaint
  theme_gdocs()+
  theme(axis.text.x = element_text(angle = 90))+
  scale_fill_jco(name = "Complaint Type")+
  guides(alpha = FALSE)
```
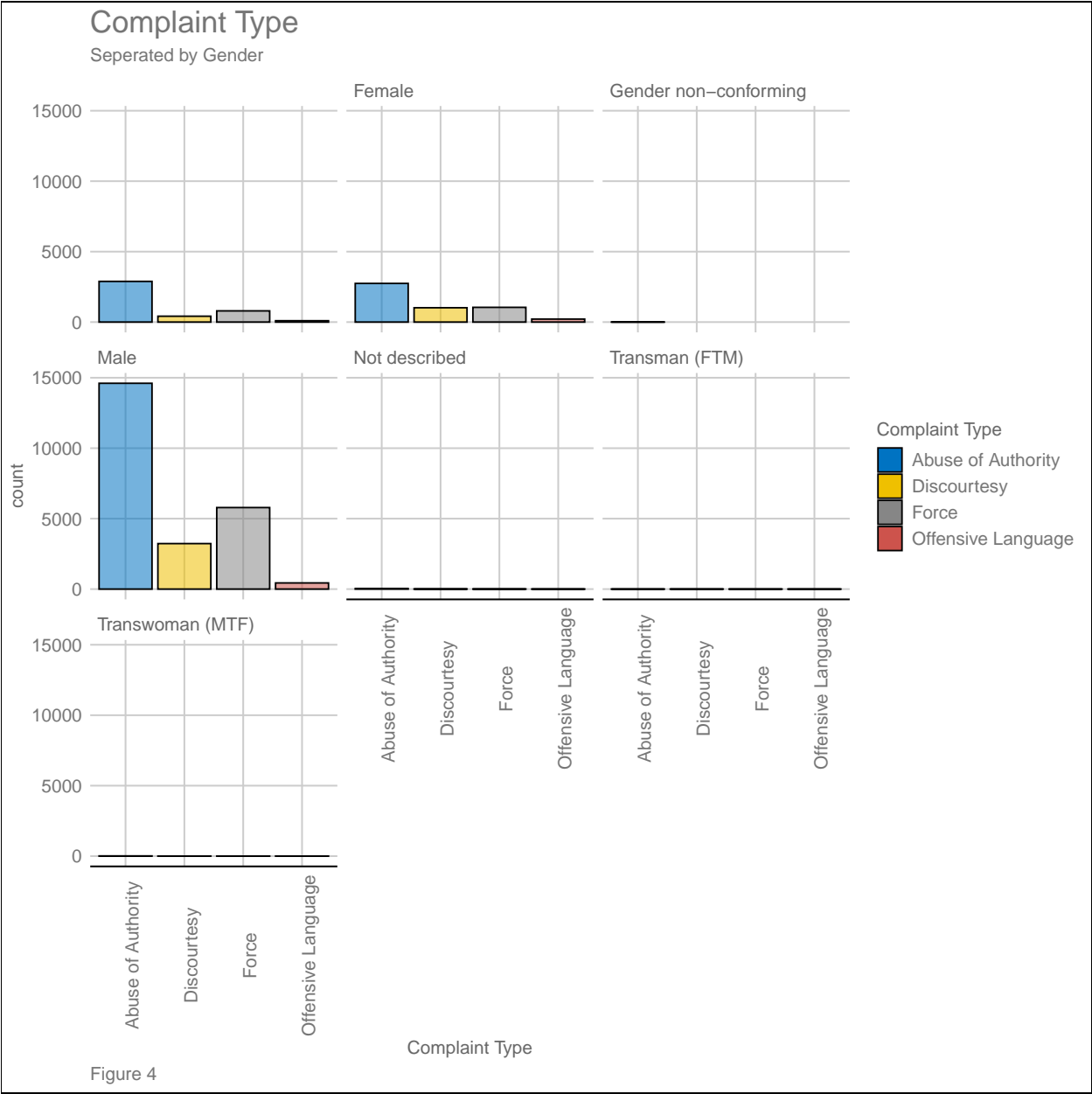
Figure 4