Colour Blind Graphs

Data this week is based of of the Tidy Tuesday dataset from the `survivoR` package. Also included is the `colorblindr` to help simulate what a graph would look like to a color blind individual. [1]

The assignment this week will have the usual grading scheme and the usual mark offs (code running off the page, not answering the question, not selecting the question) in addition to several color blind guest graders. There will be three graders who will assign 1 or 0 points based on whether or not they can read the graphs color scheme at first glance.

---

[1] *Credit to Taylor Blair for this mini-assignment.*

1. **My doctor told me I was colour blind today.**

That news really came out of the yellow.

This week you'll take on the role of a data scientist at buzzfeed who has been asked to draft a graph. Your manager has an idea for a graph they think will blow the socks off of users taking Here's How Long You Would Last On "Survivor" quizzes. You mock up the graph below (it's missing a few cities but it'll do for now). The graph should contain the following features:

   a. Scale that represents the number of winners (sole survivor and runner-up) from a given state while making it a 4-level categorical variable.
   b. Points for capital cities.
   c. Points for cities that show winners come from.

Your manager is impressed and wants you to present it to upper management to get a thumbs up before posting it on the front page! As you are about to email Gilad Lotan and the other heads of data science your boss warns you that the board is colour blind! You quickly pull up colour brewer and ggplot documentation

```r
### [BEGIN] - WORK ON YOUR DATA WRANGLING HERE
# wrangle the state information
players_by_state <- castaways %>%
  mutate(state = gsub("D.C.","District of Columbia", state),
         region = tolower(state)) %>% # cleans states
  group_by(region) %>%
  summarise(count = n()) %>% #gets counts of rows
  filter(!grepl("ontario|quebec",region)) # removes non american states

players_by_state <- players_by_state %>%
  rbind(tibble(region = setdiff(states$region,
                                players_by_state$region), # gets missing states
               count = 0)) %>%
  arrange(region) # adds on missing states

# wrangle the cities information
player_cities <- castaways %>%
  inner_join(us.cities %>%
             mutate(city=word(us.cities$name, 1))) %>%
  mutate(result = ifelse(grepl("voted out", result, fixed=T),
                         "Voted Out", result), # checks if voted out
         result = ifelse(result %in% c("Voted Out", "Runner-up", "Sole Survivor"),
                         result, "Other"),
         results = factor(result,
                          levels = c("Sole Survivor", "Runner-up", "Voted Out", "Other")))
### [END] - WORK ON YOUR DATA WRANGLING HERE
```

```
### [BEGIN] - MODIFY GGRAPH PIPELINE HERE
ggplot(data=player_cities,
              aes(x=long, y=lat, color = result)) +
  geom_polygon(data= merge(states, players_by_state),
              aes(x = long, y = lat, fill = count, group = group),
              color = "white") +
  geom_point() + # MIGHT WANT TO EDIT THIS
  scale_fill_continuous(name = "Number of Players by State") + # TRY EDITING HERE
  labs(title = "Where do Survivor Contestents live across America",
       caption = "*Only cities with more than 40,000 people are shown",
       colour = "Outcome") +
  theme(legend.position='bottom', # Cleans up the graph
        legend.box="vertical", legend.margin=margin(),
        axis.text = element_blank(), axis.line = element_blank(),
        axis.ticks = element_blank(), axis.title = element_blank(),
        panel.border = element_blank(), panel.grid = element_blank(),
        panel.background = element_blank()) +
### [END] - MODIFY GGRAPH PIPELINE HERE
  coord_fixed(1.3)
```

Where do Survivor Contestents live across America
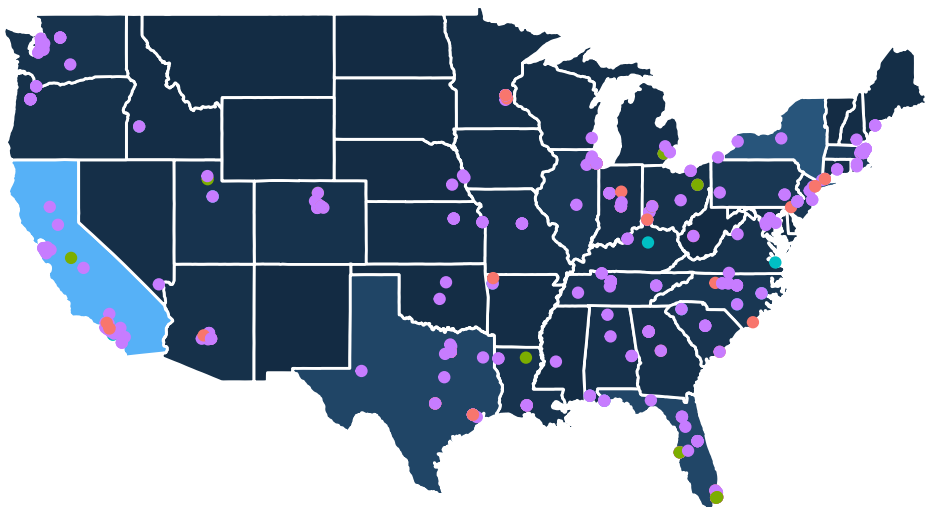
Number of Players by State

Outcome   ● Other   ● Runner−up   ● Sole Survivor   ● Voted Out

*Only cities with more than 40,000 people are shown

2. **Ad-Myer ~~Briggs~~ the pretty colors.**

Astrology for psychology

In addition to submitting your graph above your manger would like a second graph that contains **at least four features** one of which must be `personality_type`. There are other datasets in the `survivoR` including: `castaway_details`, `confessionals`, `hidden_idols`, and more! *Check out the github repo for the full list*

Any graph you make is fair game as long as it contains the minimum number of features.

    a. Chart your own course

R Pirates

Make a graph! Don't worry about making it color blind friendly ~~That's the next problem~~, just make something with features and an outline.

```
### [BEGIN] - MODIFY GGRAPH PIPELINE HERE

### [END] - MODIFY GGRAPH PIPELINE HERE
```

    b. Colour me tickled

Write your own punch line for an additional 0.1 points: [**PUNCHLINE HERE**]

Using the graph you made in the problem above (*Cough cough copy paste*), make it colour blind friendly.

It is strongly recommended you use `colorblindr`.

```
### [BEGIN] - MODIFY GGRAPH PIPELINE HERE

### [END] - MODIFY GGRAPH PIPELINE HERE
```