

**Team #: 4**

**Your name:** Julia Campbell

**Instructions:**

First, make a copy of this document (go to File then Make a copy).

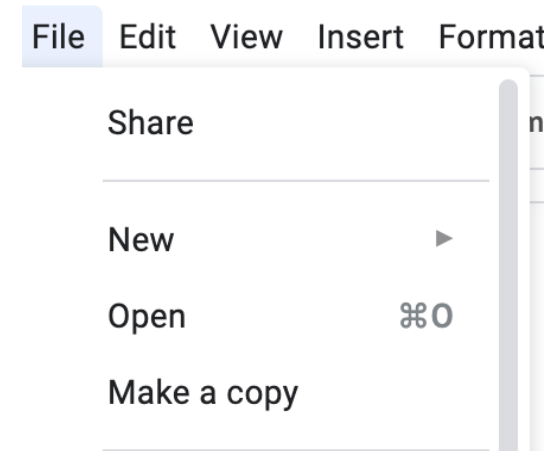
Name the document “project3 submission [ your name here ]”

This is an individual assignment. You will be making 6 graphs in total. Two bar graphs, two scatterplot graphs, and two maps. You will be comparing the weighted to the unweighted averages in the bar graphs and scatterplot graphs. You will be making a map of the county poverty rates and a map of the county bad air days.

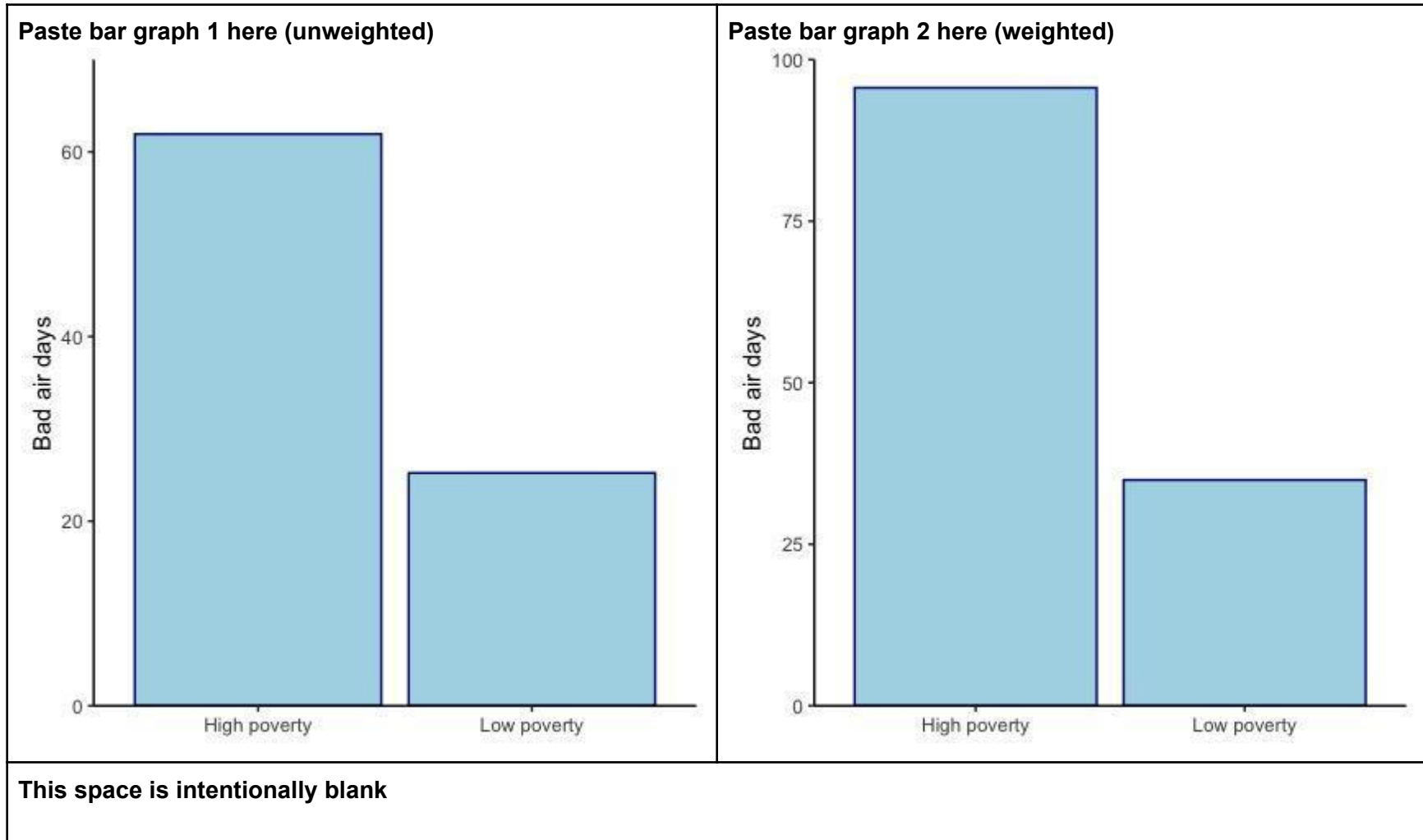
**Learning goals:** Visualize the importance of weighting. Mapping.

**CCLE:** When you are done, upload the link on CCLE.

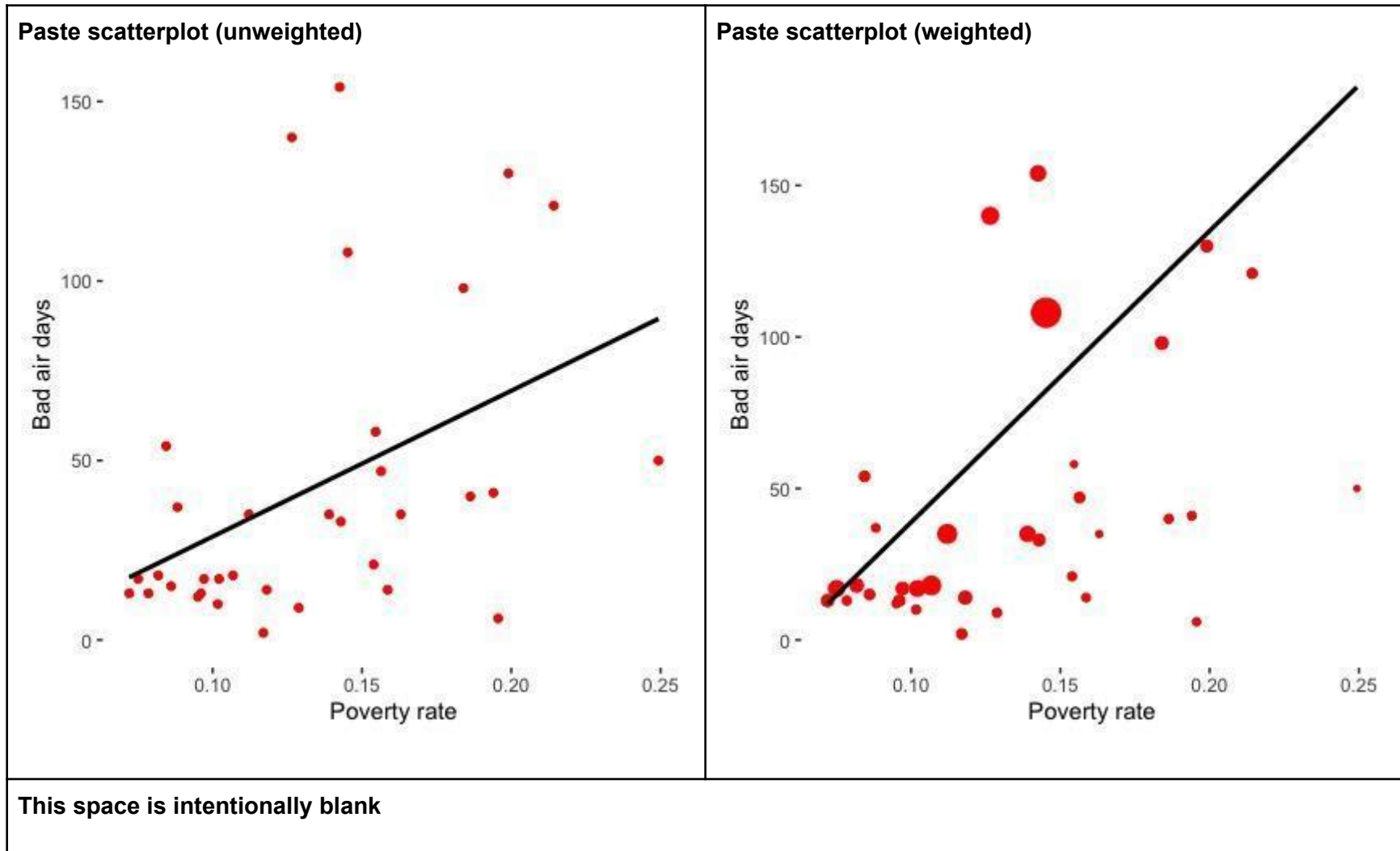
Please list your team number at the top!



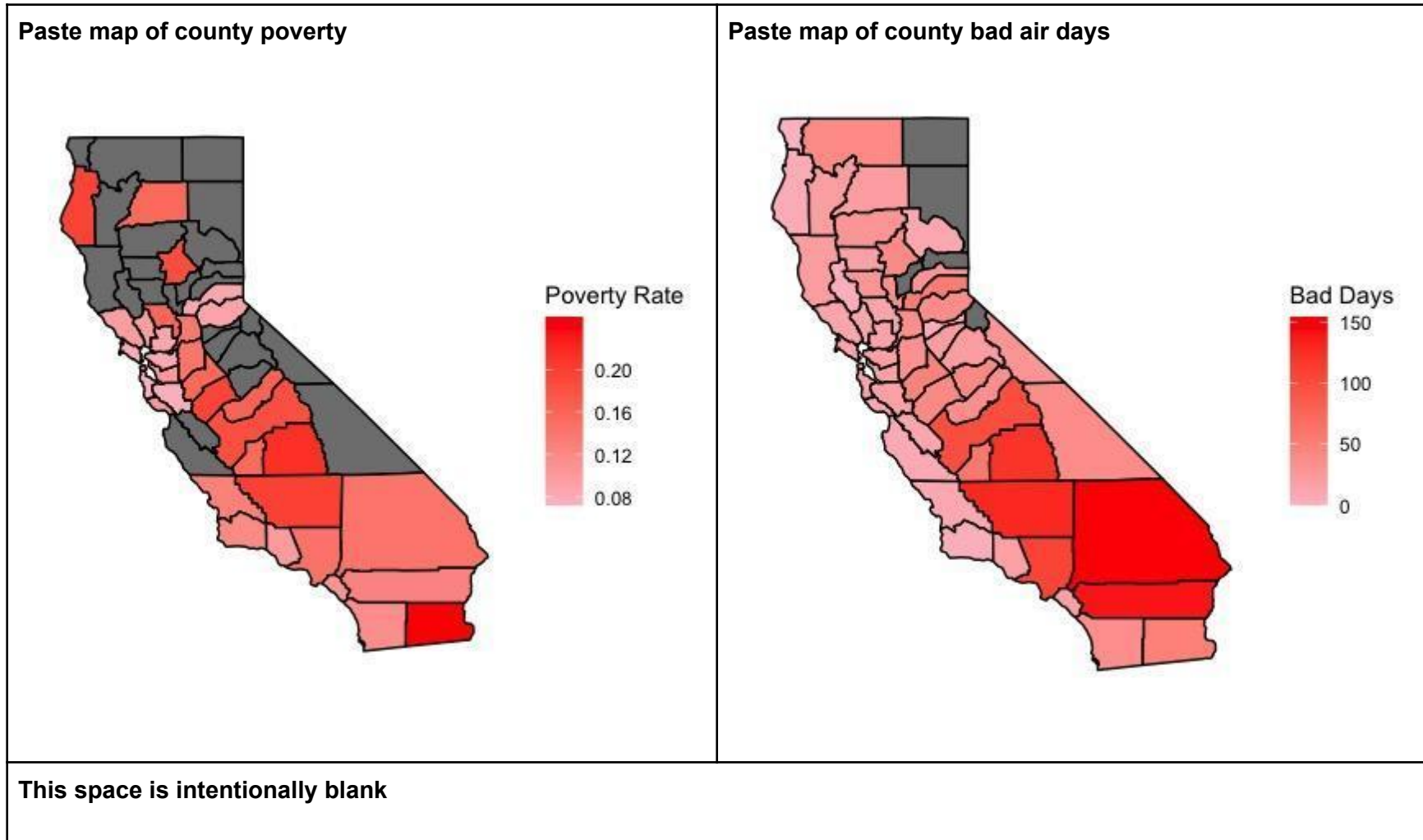
What you changed from bar graph 1 to graph 2: Changed bad\_days to bad\_days\_wgt and y scale limit to 100



What you changed from scatterplot graph 1 to graph 2: Added size=pop and weight=pop



What you changed from map graph 1 to graph 2: Changed fill=povrate100 to fill=bad\_days



**Copy and paste your R script code here.**

```
# LOAD LIBRARIES
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(readr)
```

```
library(lubridate)
```

```
library(janitor)
```

```
library(maps)
```

```
library(mapdata)
```

```
library(ggmap)
```

```
# JOIN WITH COUNTY NAMES
```

```
fips_names_california <- read_csv("~/Documents/big_enviro/data/fips_names/fips_names_california.csv")
```

```
ipums_povrate_california_2018 <- read_csv("~/Documents/big_enviro/data/ipums/ipums_povrate_california_2018.csv")
```

```
# JOIN THE NAMES
```

```
ipums_clean <- full_join(ipums_povrate_california_2018, fips_names_california, by = "countyfip")
```

```
# IMPORT DATA
```

```
annual_aqi_by_county_2018 <- read_csv("~/Documents/big_enviro/data/aqi/annual_aqi_by_county_2018.csv")
```

```
# FIX NAMES
```

```
annual_aqi_by_county_2018 <- clean_names(annual_aqi_by_county_2018)
```

```
annual_aqi_by_county_2018 <- mutate(annual_aqi_by_county_2018, state=tolower(state), county=tolower(county))
```

```
# FILTER CALIFORNIA
```

```
epa_california_2018 <- filter(annual_aqi_by_county_2018, state == 'california')
```

```
View(epa_california_2018)
```

```
# RENAME
```

```
epa_california_2018 <- rename(epa_california_2018, county_name=county)
```

```

# JOIN
final_ipums_epa <- full_join(epa_california_2018, ipums_names, by = "county_name")

# DROP UNNECESSARY VARIABLES
final_ipums_epa <- select(final_ipums_epa, -"state", -"state_name")
View(final_ipums_epa)

# BAD DAYS VARIABLE
final_ipums_epa <- mutate(final_ipums_epa,
bad_days=unhealthy_for_sensitive_groups_days+unhealthy_days+very_unhealthy_days+hazardous_days)

# MAP BORDERS
map_us_counties <- map_data("county")
map_california_counties <- filter(map_us_counties, region == "california")

# RENAME VARIABLE
map_california_counties <- rename(map_california_counties, state=region, county_name=subregion)

# JOIN W IPUMS / EPA DATA
final_map <- full_join(final_ipums_epa, map_california_counties, by = "county_name")

# POVERTY RATE MAP
ggplot(final_map, aes(y=lat, x=long, group=group, fill=povrate100)) +
  geom_polygon(color="black") +
  coord_fixed(1.3) +
  theme(panel.background = element_rect(fill="white"),
        axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text.y=element_blank(),

```

```

axis.ticks.y=element_blank() ) +
scale_fill_gradient(name = "Poverty Rate", low="pink", high="red", breaks = seq(0.08, 0.20, 0.04))
ggsave("~/Documents/big_enviro/output/tutorials/map_poverty_california_2018.pdf")

```

#### # BAD DAYS MAP

```

ggplot(final_map, aes(y=lat, x=long, group=group, fill=bad_days)) +
  geom_polygon(color="black") +
  coord_fixed(1.3) +
  theme(panel.background = element_rect(fill="white"),
        axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank() ) +
  scale_fill_gradient(name = "Bad Days", low="pink", high="red")
ggsave("~/Documents/big_enviro/output/tutorials/map_bad_days_california_2018.pdf")

```

#### # UNWEIGHTED SCATTER

```

ggplot(final_ipums_epa, aes(y=bad_days, x=povrate100)) +
  geom_point(color="red") +
  geom_smooth(method=lm, se=FALSE, color="black") +
  labs(y = "Bad air days", x = "Poverty rate", color="black") +
  theme(panel.background = element_rect(fill="white"), legend.position = "none")
ggsave("~/Documents/big_enviro/output/tutorials/scat_bad_days_povrate100_unwgt_california_2018.pdf")

```

#### # WEIGHTED SCATTER

```

ggplot(final_ipums_epa, aes(y=bad_days, x=povrate100, size=pop, weight=pop)) +
  geom_point(color="red") +
  geom_smooth(method=lm, se=FALSE, color="black") +
  labs(y = "Bad air days", x = "Poverty rate", color="black") +
  theme(panel.background = element_rect(fill="white"), legend.position = "none")

```

```

ggsave("~/Documents/big_enviro/output/tutorials/scat_bad_days_povrate100_wgt_california_2018.pdf")

# CALCULATE MEDIAN
pov_median <- median(final_ipums_epa$povrate100, na.rm = TRUE)

# GROUP
final_ipums_epa <- mutate(final_ipums_epa, povgroup=ifelse(povrate100>pov_median, "High poverty", "Low poverty"))

# GROUP BY
final_ipums_epa <- group_by(final_ipums_epa, povgroup)

# CALCULATE MEANS BY GROUP
final_ipums_epa <- mutate(final_ipums_epa, bad_days_wgt=bad_days)
final_group_sum <- summarise(final_ipums_epa, bad_days=mean(bad_days, na.rm = TRUE),
bad_days_wgt=weighted.mean(bad_days_wgt, pop, na.rm = TRUE))

# DROP MISSING OBSERVATIONS
final_group_sum <- na.omit(final_group_sum)

# UNWEIGHTED BAR
ggplot(final_group_sum, aes(y=bad_days, x=povgroup)) +
  geom_bar(stat="identity", color="dark blue", fill="light blue") +
  labs(y = "Bad air days") +
  scale_y_continuous(expand = c(0,0), limits = c(0,70)) +
  theme(panel.background = element_rect(fill="white"),
        legend.position = "none",
        axis.title.x = element_blank(),
        axis.line = element_line(colour = "black"))
ggsave("~/Documents/big_enviro/output/tutorials/bar_bad_days_povrate100_unwgt_california_2018.pdf")

# WEIGHTED BAR
ggplot(final_group_sum, aes(y=bad_days_wgt, x=povgroup)) +

```



```
geom_bar(stat="identity", color="dark blue", fill="light blue") +  
labs(y = "Bad air days") +  
scale_y_continuous(expand = c(0,0), limits = c(0,100)) +  
theme(panel.background = element_rect(fill="white"),  
      legend.position = "none",  
      axis.title.x = element_blank(),  
      axis.line = element_line(colour = "black"))  
ggsave("~/Documents/big_enviro/output/tutorials/bar_bad_days_povrate100_wgt_california_2018.pdf")
```