

Lab 2: Descriptive Statistics and Basic Mapping

GIS III: Geocomputation. University of Chicago, Spring 2020

Susan Paykin

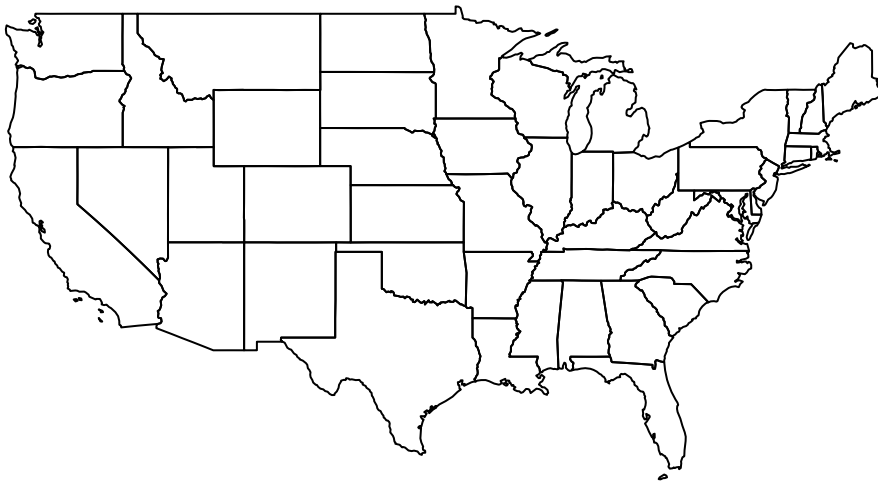
Load and explore spatial dataset on United States population

Data source: spData package

```
names(us_states)
```

```
## [1] "GEOID"      "NAME"      "REGION"    "AREA"      "total_pop_10"  
## [6] "total_pop_15" "geometry"
```

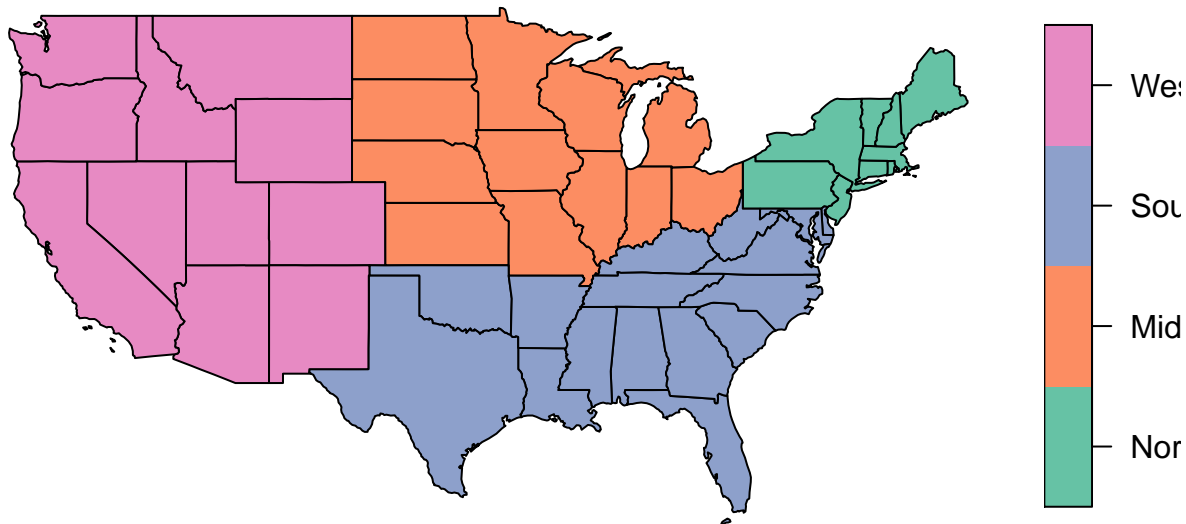
```
plot(us_states["geometry"])
```



Summarize poulation by region and number of states per region

```
# Exploratory plot of regions  
plot(us_states["REGION"])
```

REGION



```
# Summarize population by region and create table
region_pop <- us_states %>%
  rename(region = REGION) %>%
  group_by(region) %>%
  summarize(pop_2015 = comma(sum(total_pop_15)), n_states = n()) %>%
  arrange(desc(pop_2015)) %>%
  st_drop_geometry()

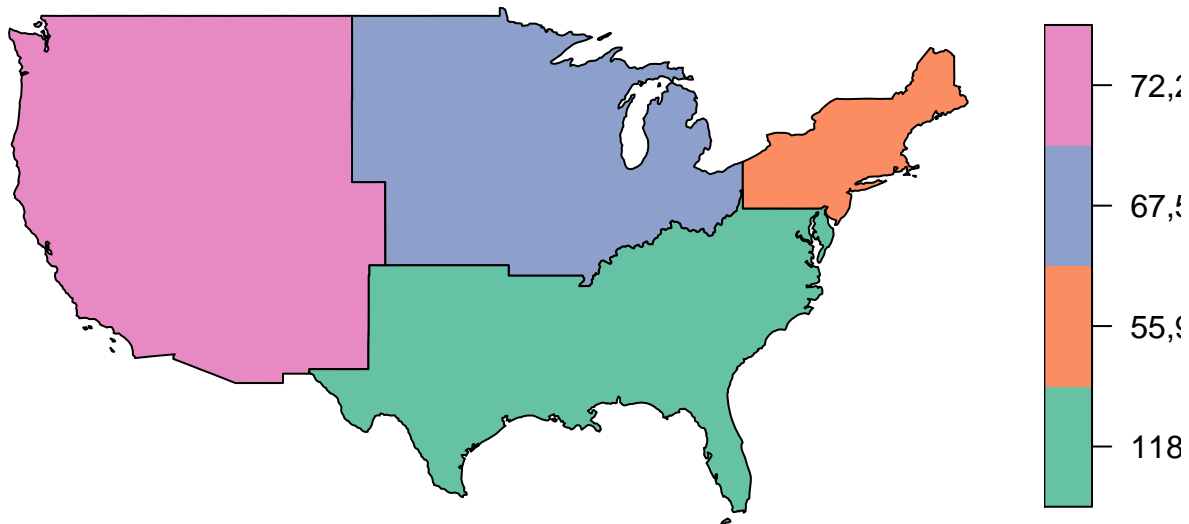
kable(region_pop)
```

region	pop_2015	n_states
West	72,264,052	11
Midwest	67,546,398	12
Northeast	55,989,520	9
South	118,575,377	17

```
# Keep geometry
region_pop_geom <- us_states %>%
  rename(region = REGION) %>%
  group_by(region) %>%
  summarize(pop_2015 = comma(sum(total_pop_15)), n_states = n()) %>%
  arrange(desc(pop_2015))

# Plot population by region
plot(region_pop_geom["pop_2015"], main = "Population by Region, 2015")
```

Population by Region, 2015



Identify ten most populous US states

Plot population, group by region and show land area (sq kilometers)

```
top10 <- us_states %>%
  select(NAME, AREA, REGION, total_pop_15) %>%
  arrange(desc(total_pop_15)) %>%
  slice(1:10)

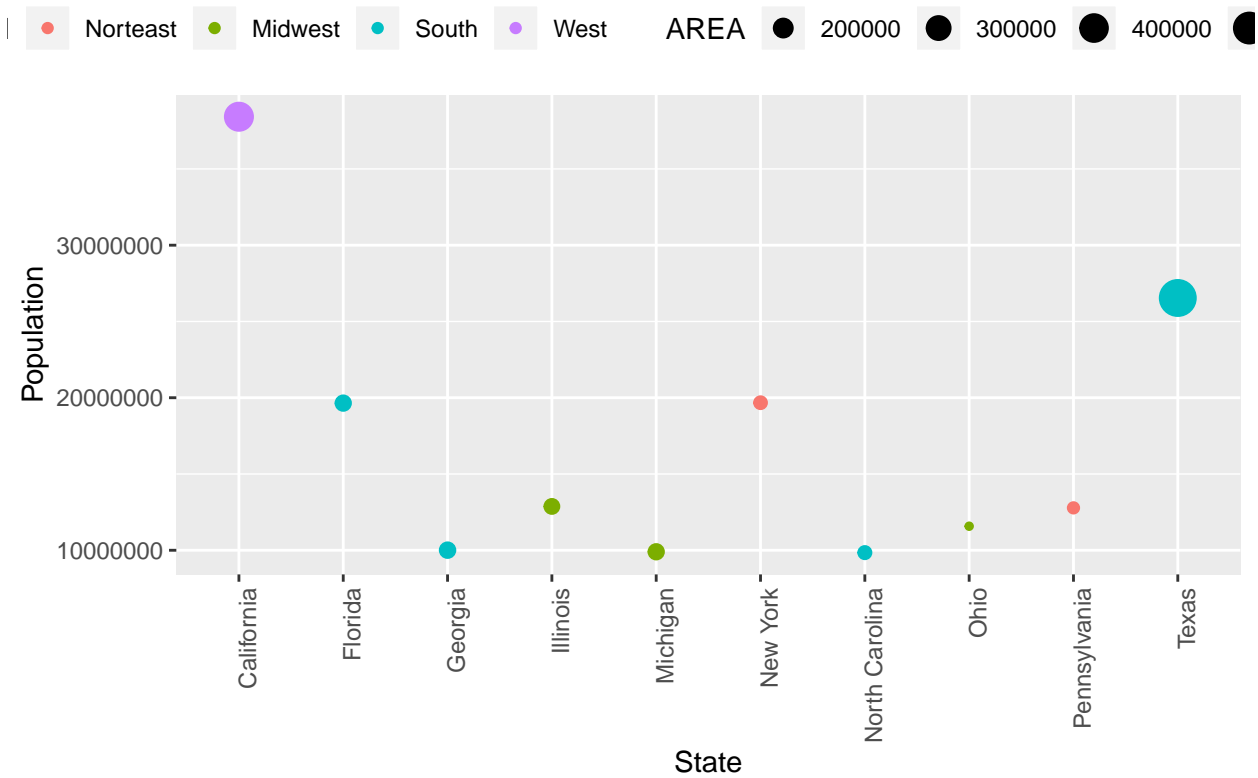
top10$AREA <- as.numeric(top10$AREA)

options(scipen = 999)

p <- ggplot(data = top10, aes(x = NAME, y = total_pop_15, size = AREA, color = REGION)) +
  geom_point() +
  labs(title = "Top 10 US State Populations", x = "State", y = "Population") +
  theme(legend.position = "top")

p + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Top 10 US State Populations



Map ten most populous US states

```
plot(st_geometry(us_states), main = "Ten Most Populous US States")
us_states %>%
  select(total_pop_15) %>%
  arrange(desc(total_pop_15)) %>%
  slice(1:10) %>%
  plot(add = TRUE, col = 'pink')
```

Ten Most Populous US States

