Homework 4 Solution

Byteflow Dynamics

10/15/2017

Strings

Create regular expressions to do following:

1. Match both "doughnut" and "donut" (you can do this in more than one ways)

```
str_view(c("doughnut", "donut"), "do(ugh)?nut")
```

```
str_view(c("doughnut", "donut"), "d(o|ough)nut")
```

2. Split the following date "10/15/2017" into each components: month, day, year.

```
date <- "10/15/2017"
split_date <- str_split(date, "/")
split_date
```

```
## [[1]]
## [1] "10"   "15"   "2017"
```

3. Transform a string "10/15/2017" into "10-15-2017"

```
str_replace_all(date, "/", "-")
```

```
## [1] "10-15-2017"
```

4. We will use the **sentences** data.

```
length(sentences)
```

```
## [1] 720
```

```
#> [1] 720
head(sentences)
```

```
## [1] "The birch canoe slid on the smooth planks."
## [2] "Glue the sheet to the dark blue background."
## [3] "It's easy to tell the depth of a well."
## [4] "These days a chicken leg is a rare dish."
## [5] "Rice is often served in round bowls."
## [6] "The juice of lemons makes fine punch."
```

```
#> [1] "The birch canoe slid on the smooth planks."
#> [2] "Glue the sheet to the dark blue background."
#> [3] "It's easy to tell the depth of a well."
#> [4] "These days a chicken leg is a rare dish."
#> [5] "Rice is often served in round bowls."
#> [6] "The juice of lemons makes fine punch."
```

- Match sentences that start with "The".

```
str_view(sentences, "^The ", match = TRUE)
```

- Match sentences that include the word 'good'.

```
str_view(sentences, "good ", match = TRUE)
```

- Match the last 3 characters of each sentence including the period.

```
str_view(sentences, "..\\.")
```

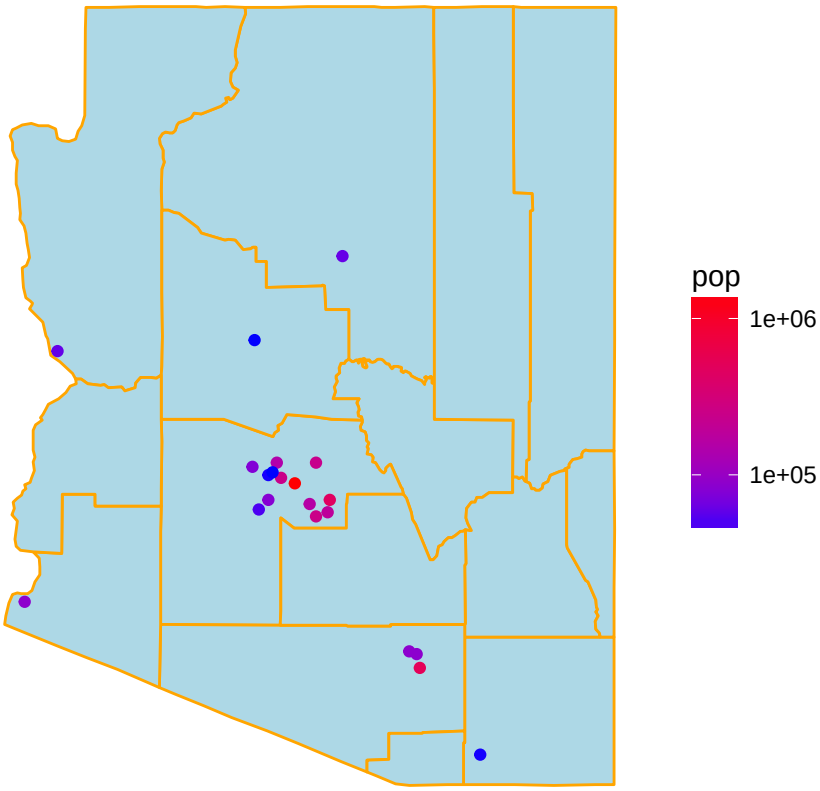
ggplot

Recreate this plot. This is similar to the NY state plot we made in class, but for Arizona.

Don't worry too much about the colors. You can play with different colors of your choice.

```
counties <- map_data("county")
az_county <- subset(counties, region == "arizona")

azcounties <- ggplot(data = az_county) +
  geom_polygon(mapping = aes(x = long, y = lat, group = group), fill = "light blue", color = "orange") +
  coord_fixed(1.3)

data(us.cities)

azcities <- subset(us.cities, country.etc == "AZ")

azcityplot <- azcounties +
  geom_point(data = azcities, mapping = aes(x = long, y = lat, color = pop)) +
  scale_colour_gradient(high = "red", low = "blue", trans = "log10")

azcityplot +
  theme_void()
```

```
## Warning in plyr::split_indices(scale_id, n): '.Random.seed' is not an
## integer vector but of type 'NULL', so ignored
```


pop
1e+06
1e+05