Data Res Graphs

Jaclyn Chiu

11/8/2020

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2
```

```
## -- Attaching packages ---------------------------------------------------------------------------------------- tidyver
se 1.3.0 --
```

```
## v ggplot2 3.3.0     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## Warning: package 'tidyr' was built under R version 4.0.2
```

```
## -- Conflicts ------------------------------------------------------------------------------------------------ tidyverse_con
flicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.0.3
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.0.2
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:dplyr':
##
##     intersect, setdiff, union
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
# explore into data 1
data1 <- read_csv("California_Fire_Incidents.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Active = col_logical(),
##   AdminUnit = col_character(),
##   CalFireIncident = col_logical(),
##   CanonicalUrl = col_character(),
##   ConditionStatement = col_character(),
##   ControlStatement = col_character(),
##   Counties = col_character(),
##   CountyIds = col_character(),
##   Extinguished = col_datetime(format = ""),
##   Featured = col_logical(),
##   Final = col_logical(),
##   FuelType = col_logical(),
##   Location = col_character(),
##   MajorIncident = col_logical(),
##   Name = col_character(),
##   Public = col_logical(),
##   SearchDescription = col_character(),
##   SearchKeywords = col_character(),
##   Started = col_datetime(format = ""),
##   Status = col_character()
##   # ... with 3 more columns
## )
```

```
## See spec(...) for full column specifications.
```

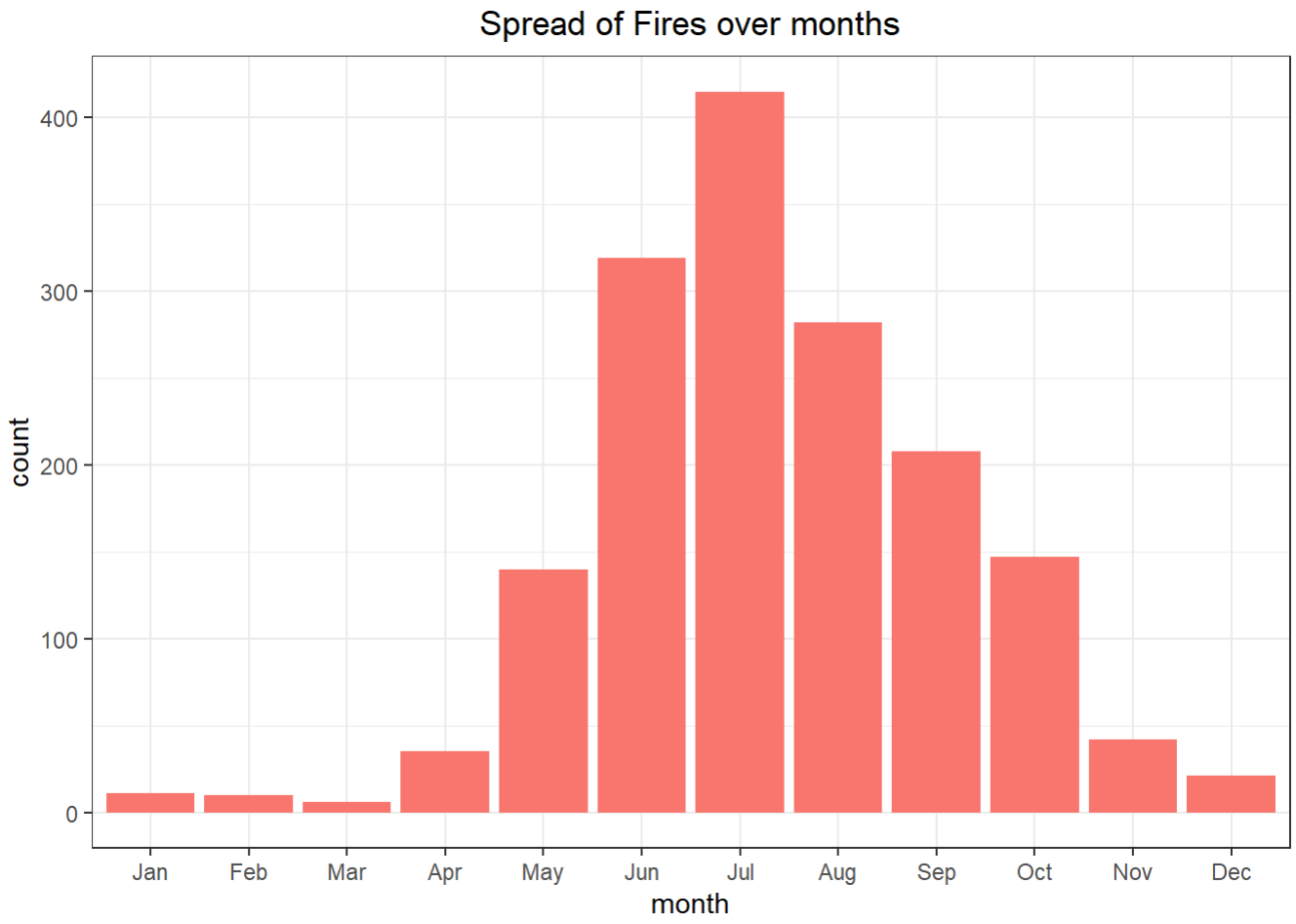
```
## Warning: 12 parsing failures.
##  row      col           expected     actual                            file
## 1387 FuelType 1/0/T/F/TRUE/FALSE Grass      'California_Fire_Incidents.csv'
## 1392 FuelType 1/0/T/F/TRUE/FALSE Grass      'California_Fire_Incidents.csv'
## 1396 FuelType 1/0/T/F/TRUE/FALSE timber     'California_Fire_Incidents.csv'
## 1397 FuelType 1/0/T/F/TRUE/FALSE Vegetation 'California_Fire_Incidents.csv'
## 1403 FuelType 1/0/T/F/TRUE/FALSE Brush      'California_Fire_Incidents.csv'
## .... ........ .................. .......... ..............................
## See problems(...) for more details.
```

```
view(data1)
dim(data1)
```

```
## [1] 1636   40
```

```
#extract month of each fire and plot data
data1$month <- month(data1$Started)

# GRAPH 1
ggplot(data1, aes(x = month, fill = "red")) + geom_bar() +
  scale_x_discrete(limits = 1:12, labels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Au
g", "Sep", "Oct", "Nov", "Dec")) +
  theme_bw() + theme(legend.position = "none", plot.title = element_text(hjust = 0.5)) +
  ggtitle("Spread of Fires over months")
```




```
#counties with most amount of fires

data1 %>% group_by(Counties) %>% summarise(n = n()) %>% arrange(desc(n))
```

```
## # A tibble: 59 x 2
##    Counties            n
##    <chr>           <int>
##  1 Riverside         146
##  2 San Diego          89
##  3 Butte              66
##  4 San Luis Obispo    64
##  5 Shasta             64
##  6 Kern               62
##  7 Fresno             57
##  8 Siskiyou           57
##  9 San Bernardino     53
## 10 Tehama             51
## # ... with 49 more rows
```

```
counties <- c("Riverside", "San Diego", "Butte", "San Luis Obispo", "Shasta")
number <- c(146, 89, 66, 64, 64)
new <- data.frame(counties = counties, number = number)

# GRAPH 2
new %>% mutate(counties = fct_reorder(counties, desc(number))) %>% ggplot(aes(x = counties, y =
 number, fill = counties)) + geom_bar(stat = "identity") + ggtitle("Counties with Most Fires") +
  theme_bw() + theme(plot.title = element_text(hjust = 0.5))
```