

Milestone 2 data cleaning

Tsz Yau Iris Chow

3/10/2021

```
rm(list = ls())  
library("dplyr")
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library("tidyr")
```

```
carbon <- read.csv("carbon-monitor-US.csv")
```

```
#remove rows with NA  
carbon <- na.omit(carbon)  
  
#remove country column  
carbon$country...group.of.countries = NULL  
  
head(carbon)
```

```
##           date sector MtCO2.per.day  
## 1 01/01/2019  Power         3.962845  
## 2 02/01/2019  Power         4.617279  
## 3 03/01/2019  Power         4.444041  
## 4 04/01/2019  Power         4.352839  
## 5 05/01/2019  Power         4.027248  
## 6 06/01/2019  Power         3.444597
```

```
#modify dates  
carbon <- carbon %>%  
  mutate(date = as.Date(date, format = "%d/%m/%Y"))  
  
carbon_date <- carbon
```

```
#remove sector column
carbon_date$sector = NULL
```

```
#combine dates into one row per day
carbon_date = aggregate(.-date, data = carbon_date, FUN = sum)
head(carbon_date)
```

```
##           date MtCO2.per.day
## 1 2019-01-01      13.15755
## 2 2019-01-02      15.67320
## 3 2019-01-03      15.41266
## 4 2019-01-04      15.28831
## 5 2019-01-05      14.40894
## 6 2019-01-06      12.92496
```

```
#Widen table by sector
carbon_sector <- carbon %>%
  pivot_wider(names_from = sector, values_from = MtCO2.per.day)
head(carbon_sector)
```

```
## # A tibble: 6 x 6
##   date      Power 'Ground Transport' Industry Residential 'Domestic Aviation'
##   <date>    <dbl>          <dbl>    <dbl>          <dbl>          <dbl>
## 1 2019-01-01  3.96            3.35      2.85            2.61            0.380
## 2 2019-01-02  4.62            4.30      3.13            3.15            0.473
## 3 2019-01-03  4.44            4.43      3.07            2.96            0.509
## 4 2019-01-04  4.35            4.60      3.04            2.80            0.496
## 5 2019-01-05  4.03            4.45      2.90            2.59            0.447
## 6 2019-01-06  3.44            3.97      2.60            2.48            0.427
```

creating graphs

```
library("ggplot2")
library("dplyr")
library("dslabs")
library("readr")
theme_set(theme_bw())
```

```
ggplot(carbon, aes(x = date, y = MtCO2.per.day, col = sector)) +
  geom_line() +
  ggtitle("Carbon Emissions in the US by Sector") +
  labs(y = "Carbon Emissions (Mt)") +
  geom_vline(xintercept = as.Date("19/01/2020", "%d/%m/%Y"), color = "red", linetype = "solid")
```

