Geolocating Wineries

```
knitr::opts_chunk$set(
                       collapse = TRUE,
                       eval = FALSE)
options(crayon.enabled = FALSE)
```

Using ggmap

```
library(ggmap)
library(tidyverse)
```

```
data_dir <- "../data"
file_name <- "winemag-data-130k-v2.csv"
path <- file.path(data_dir, file_name)
```

```
Wine <- read_csv(path,
                 col_types = cols(
                     X1 = col_double(),
                     country = col_character(),
                     description = col_character(),
                     designation = col_character(),
                     points = col_double(),
                     price = col_double(),
                     province = col_character(),
                     region_1 = col_character(),
                     region_2 = col_character(),
                     taster_name = col_character(),
                     taster_twitter_handle = col_character(),
                     title = col_character(),
                     variety = col_character(),
                     winery = col_character()),
                 progress = FALSE
                 ) %>%
    rename(id = X1) %>%
    mutate(id = id + 1)
```

The `mutate_geocoded` function from `ggmap` would be great if it actually had some error handling. Instead, we will form our own function.

```
subset <- Wine %>%
    count(winery, country) %>%
    mutate(address = paste0(winery, " ", country)) %>%
    sample_n(10)
```

```
geocode_robustly <-
    possibly(
        insistently(
            quietly(geocode),
            rate = rate_delay(0.1, max_times = 2)),
        otherwise = list(result = tibble(lon = NA_real_, lat = NA_real_))
    )
```

```
locations <- subset %>%
    pull(address) %>%
    map_dfr(~ geocode_robustly(.x)$result)

subset %>% bind_cols(locations)
```

And now we geocode all of the dataset.

```
# 1. Add address column to geocode
Wine <- Wine %>%
    mutate(address = paste0(winery, " ", country))

# 2. Get unique addresses
Addresses <- Wine %>%
    count(address) %>%
    select(-n)

# 3. Geocode unique addresses
Locations <- Addresses %>%
    pull(address) %>%
    map_dfr(~ geocode_robustly(.x)$result)

# 4. Bind location info to addresses
Addresses <- Addresses %>%
    bind_cols(Locations)

# 5. Join into original dataset
Geocoded_Wine <- Wine %>%
    left_join(Addresses, by = "address")
options(tinytex.verbose = TRUE)
# 4. Save geocoded dataset
file_name <- "geocoded.csv"
Geocoded_Wine %>% write_csv(path = file.path(data_dir, file_name))
```

Verify

```
Wine <- read_csv("../data/geocoded.csv") %>%
    glimpse()
## Observations: 129,971
## Variables: 17
## $ id                    <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1...
## $ country               <chr> "Italy", "Portugal", "US", "US", "US", "...
## $ description           <chr> "Aromas include tropical fruit, broom, b...
## $ designation           <chr> "Vulkà Bianco", "Avidagos", NA, "Reserve...
## $ points                <dbl> 87, 87, 87, 87, 87, 87, 87, 87, 87, 87, ...
## $ price                 <dbl> NA, 15, 14, 13, 65, 15, 16, 24, 12, 27, ...
## $ province              <chr> "Sicily & Sardinia", "Douro", "Oregon", ...
## $ region_1              <chr> "Etna", NA, "Willamette Valley", "Lake M...
## $ region_2              <chr> NA, NA, "Willamette Valley", NA, "Willam...
## $ taster_name           <chr> "Kerin O'Keefe", "Roger Voss", "Paul Gre...
## $ taster_twitter_handle <chr> "@kerinokeefe", "@vossroger", "@paulgwin...
## $ title                 <chr> "Nicosia 2013 Vulkà Bianco  (Etna)", "Qu...
## $ variety               <chr> "White Blend", "Portuguese Red", "Pinot ...
## $ winery                <chr> "Nicosia", "Quinta dos Avidagos", "Rains...
```

```
## $ address               <chr> "Nicosia Italy", "Quinta dos Avidagos Po...
## $ lon                   <dbl> 14.395278, -7.276971, -95.712891, -85.89...
## $ lat                   <dbl> 37.74692, 41.38793, 37.09024, 42.21225, ...
```