# Getting Started With the COVID-19 Data

Michael Hahsler

2021-01-29

- COVID 19 Cases in Texas Counties
  - Are there many counties with many cases?
  - Relationship between cases and deaths
  - Does death per case depend on population?
  - What variables are correlated?
  - Plot as a map
- Look at Dallas County over time. Are we flattening the curve?

I use mostly tidyverse and ggplot. The focus of the analysis is Texas and Dallas County, TX.

```
library("tidyverse")
```

```
## ── Attaching packages ─────────────────────────────────────────────────────────────────────────────── ti
dyverse 1.3.0 ──
```

```
## ✓ ggplot2 3.3.2     ✓ purrr   0.3.4
## ✓ tibble  3.0.3     ✓ dplyr   1.0.0
## ✓ tidyr   1.1.0     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.5.0
```

```
## ── Conflicts ────────────────────────────────────────────────────────────────────────────────── tidyvers
e_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("ggplot2")
library("ggrepel")
library("ggcorrplot")
library("DT")
```

# COVID 19 Cases in Texas Counties

Data is for May 28, 2020 and was obtained from

https://console.cloud.google.com/marketplace/details/usafacts-public-data/covid19-us-cases?filter=solution-type:dataset&filter=category:covid19&id=3eaff9c5-fbaf-47bb-a441-89db1e1395ab (https://console.cloud.google.com/marketplace/details/usafacts-public-data/covid19-us-cases?filter=solution-type:dataset&filter=category:covid19&id=3eaff9c5-fbaf-47bb-a441-89db1e1395ab)

using the query

```
SELECT *
FROM `bigquery-public-data.covid19_usafacts.summary` covid19
JOIN `bigquery-public-data.census_bureau_acs.county_2017_5yr` acs
ON covid19.county_fips_code = acs.geo_id
WHERE date = DATE_SUB(CURRENT_DATE(), INTERVAL 7 day)
```

```
cases <- read_csv("COVID-19_cases_plus_census.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   county_fips_code = col_character(),
##   county_name = col_character(),
##   state = col_character(),
##   state_fips_code = col_character(),
##   date = col_date(format = ""),
##   geo_id = col_character(),
##   pop_5_years_over = col_logical(),
##   speak_only_english_at_home = col_logical(),
##   speak_spanish_at_home = col_logical(),
##   speak_spanish_at_home_low_english = col_logical(),
##   pop_15_and_over = col_logical(),
##   pop_never_married = col_logical(),
##   pop_now_married = col_logical(),
##   pop_separated = col_logical(),
##   pop_widowed = col_logical(),
##   pop_divorced = col_logical()
## )
```

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

```
cases
```

```
## # A tibble: 3,142 x 259
##    county_fips_code county_name state state_fips_code date       confirmed_cases
##    <chr>            <chr>       <chr> <chr>           <date>               <dbl>
##  1 50009            Essex Coun… VT    50              2021-01-19             111
##  2 50007            Chittenden… VT    50              2021-01-19            3636
##  3 10001            Kent County DE    10              2021-01-19           11548
##  4 44009            Washington… RI    44              2021-01-19            5521
##  5 33001            Belknap Co… NH    33              2021-01-19            2496
##  6 44005            Newport Co… RI    44              2021-01-19            3578
##  7 50015            Lamoille C… VT    50              2021-01-19             312
##  8 09013            Tolland Co… CT    09              2021-01-19            6255
##  9 50001            Addison Co… VT    50              2021-01-19             527
## 10 50005            Caledonia … VT    50              2021-01-19             307
## # … with 3,132 more rows, and 253 more variables: deaths <dbl>, geo_id <chr>,
## #   nonfamily_households <dbl>, family_households <dbl>,
## #   median_year_structure_built <dbl>, rent_burden_not_computed <dbl>,
## #   rent_over_50_percent <dbl>, rent_40_to_50_percent <dbl>,
## #   rent_35_to_40_percent <dbl>, rent_30_to_35_percent <dbl>,
## #   rent_25_to_30_percent <dbl>, rent_20_to_25_percent <dbl>,
## #   rent_15_to_20_percent <dbl>, rent_10_to_15_percent <dbl>,
## #   rent_under_10_percent <dbl>, total_pop <dbl>, male_pop <dbl>,
## #   female_pop <dbl>, median_age <dbl>, white_pop <dbl>, black_pop <dbl>,
## #   asian_pop <dbl>, hispanic_pop <dbl>, amerindian_pop <dbl>,
## #   other_race_pop <dbl>, two_or_more_races_pop <dbl>, not_hispanic_pop <dbl>,
## #   commuters_by_public_transportation <dbl>, households <dbl>,
## #   median_income <dbl>, income_per_capita <dbl>, housing_units <dbl>,
## #   vacant_housing_units <dbl>, vacant_housing_units_for_rent <dbl>,
## #   vacant_housing_units_for_sale <dbl>, median_rent <dbl>,
## #   percent_income_spent_on_rent <dbl>, owner_occupied_housing_units <dbl>,
## #   million_dollar_housing_units <dbl>, mortgaged_housing_units <dbl>,
## #   families_with_young_children <dbl>,
## #   two_parent_families_with_young_children <dbl>,
## #   two_parents_in_labor_force_families_with_young_children <dbl>,
## #   two_parents_father_in_labor_force_families_with_young_children <dbl>,
## #   two_parents_mother_in_labor_force_families_with_young_children <dbl>,
## #   two_parents_not_in_labor_force_families_with_young_children <dbl>,
## #   one_parent_families_with_young_children <dbl>,
## #   father_one_parent_families_with_young_children <dbl>,
## #   father_in_labor_force_one_parent_families_with_young_children <dbl>,
## #   commute_10_14_mins <dbl>, commute_15_19_mins <dbl>,
## #   commute_20_24_mins <dbl>, commute_25_29_mins <dbl>,
## #   commute_30_34_mins <dbl>, commute_45_59_mins <dbl>,
## #   aggregate_travel_time_to_work <dbl>, income_less_10000 <dbl>,
## #   income_10000_14999 <dbl>, income_15000_19999 <dbl>,
## #   income_20000_24999 <dbl>, income_25000_29999 <dbl>,
## #   income_30000_34999 <dbl>, income_35000_39999 <dbl>,
## #   income_40000_44999 <dbl>, income_45000_49999 <dbl>,
## #   income_50000_59999 <dbl>, income_60000_74999 <dbl>,
## #   income_75000_99999 <dbl>, income_100000_124999 <dbl>,
## #   income_125000_149999 <dbl>, income_150000_199999 <dbl>,
## #   income_200000_or_more <dbl>,
## #   renter_occupied_housing_units_paying_cash_median_gross_rent <dbl>,
## #   owner_occupied_housing_units_lower_value_quartile <dbl>,
## #   owner_occupied_housing_units_median_value <dbl>,
## #   owner_occupied_housing_units_upper_value_quartile <dbl>,
## #   married_households <dbl>, occupied_housing_units <dbl>,
## #   housing_units_renter_occupied <dbl>, dwellings_1_units_detached <dbl>,
## #   dwellings_1_units_attached <dbl>, dwellings_2_units <dbl>,
## #   dwellings_3_to_4_units <dbl>, dwellings_5_to_9_units <dbl>,
## #   dwellings_10_to_19_units <dbl>, dwellings_20_to_49_units <dbl>,
## #   dwellings_50_or_more_units <dbl>, mobile_homes <dbl>,
## #   housing_built_2005_or_later <dbl>, housing_built_2000_to_2004 <dbl>,
## #   housing_built_1939_or_earlier <dbl>, male_under_5 <dbl>, male_5_to_9 <dbl>,
## #   male_10_to_14 <dbl>, male_15_to_17 <dbl>, male_18_to_19 <dbl>,
## #   male_20 <dbl>, male_21 <dbl>, male_22_to_24 <dbl>, male_25_to_29 <dbl>, …
```

Make character factors for analysis

```
cases <- cases %>% mutate_if(is.character, factor)
dim(cases)
```

```
## [1] 3142  259
```

Filter Texas

```
cases_TX <- cases %>% filter(state == "TX")
dim(cases_TX)
```

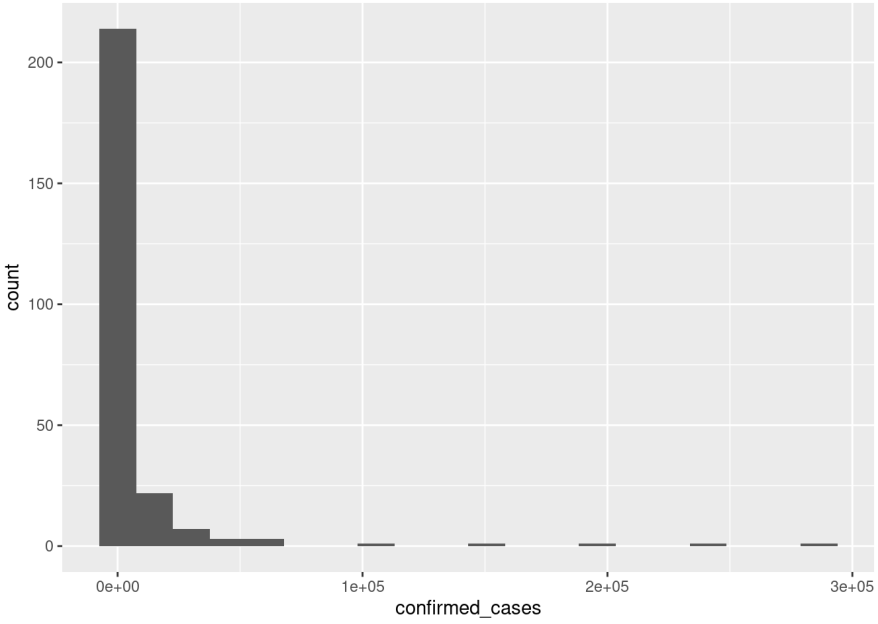
```
## [1] 254 259
```

```
summary(cases_TX[,1:10])
```

```
##  county_fips_code           county_name      state     state_fips_code
##  48001  :  1      Anderson County :  1   TX     :254   48     :254
##  48003  :  1      Andrews County  :  1   AK     :  0   01     :  0
##  48005  :  1      Angelina County :  1   AL     :  0   02     :  0
##  48007  :  1      Aransas County  :  1   AR     :  0   04     :  0
##  48009  :  1      Archer County   :  1   AZ     :  0   05     :  0
##  48011  :  1      Armstrong County:  1   CA     :  0   06     :  0
##  (Other):248      (Other)         :248   (Other):  0   (Other):  0
##       date            confirmed_cases      deaths            geo_id
##  Min.   :2021-01-19   Min.   :     1   Min.   :   0.00   48001  :  1
##  1st Qu.:2021-01-19   1st Qu.:   487   1st Qu.:  13.00   48003  :  1
##  Median :2021-01-19   Median :  1310   Median :  30.00   48005  :  1
##  Mean   :2021-01-19   Mean   :  8419   Mean   : 127.48   48007  :  1
##  3rd Qu.:2021-01-19   3rd Qu.:  3502   3rd Qu.:  78.75   48009  :  1
##  Max.   :2021-01-19   Max.   :286356   Max.   :3825.00   48011  :  1
##                                                          (Other):248
##  nonfamily_households family_households
##  Min.   :    12       Min.   :     19
##  1st Qu.:   674       1st Qu.:   1618
##  Median :  1883       Median :   4474
##  Mean   : 11300       Mean   :  25828
##  3rd Qu.:  4683       3rd Qu.:  12366
##  Max.   :496164       Max.   :1066649
##
```

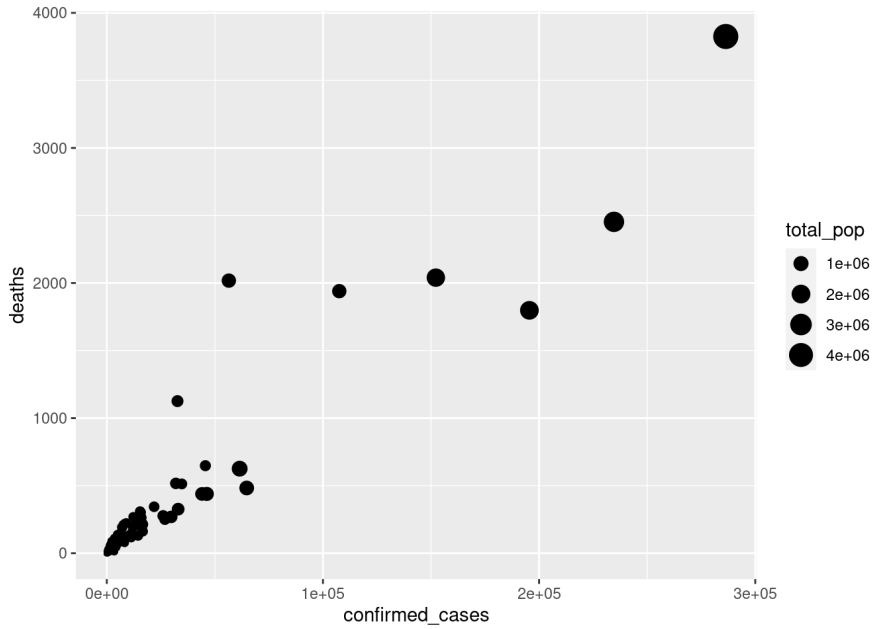
## Are there many counties with many cases?

```
ggplot(cases_TX, mapping = aes(confirmed_cases)) + geom_histogram(bins = 20)
```



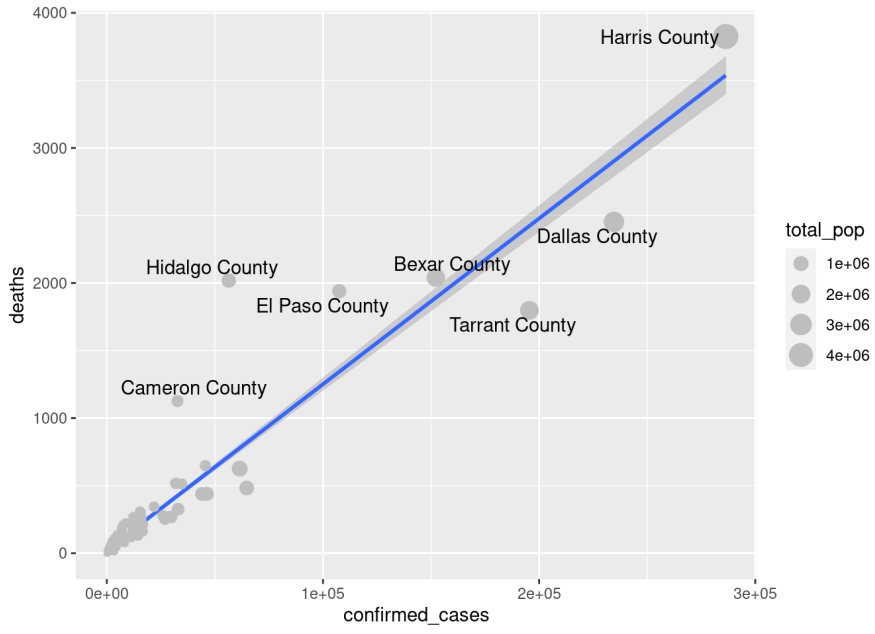

## Relationship between cases and deaths

```
ggplot(cases_TX, mapping = aes(x = confirmed_cases, y = deaths, size = total_pop)) + geom_point()
```




```
ggplot(cases_TX, mapping = aes(x = confirmed_cases, y = deaths, label = county_name)) +
  geom_smooth(method = lm) +
  geom_point(mapping = aes(size = total_pop), color = "grey") +
  geom_text_repel(data = subset(cases_TX, deaths >= 1000))
```

```
## `geom_smooth()` using formula 'y ~ x'
```




Calculate rates (per 1000 people)i

```
cases_TX_select <- cases_TX %>% filter(confirmed_cases > 100) %>%
  arrange(desc(confirmed_cases)) %>%
  select(county_name, confirmed_cases, deaths, total_pop, median_income)
cases_TX_select <- cases_TX_select %>% mutate(
  cases_per_1000 = confirmed_cases/total_pop*1000,
  deaths_per_1000 = deaths/total_pop*1000,
  death_per_case = deaths/confirmed_cases)

head(cases_TX_select)
```

```
## # A tibble: 6 x 8
##   county_name confirmed_cases deaths total_pop median_income cases_per_1000
##   <fct>                 <dbl>  <dbl>     <dbl>         <dbl>          <dbl>
## 1 Harris Cou…           286356   3825   4525519         57791           63.3
## 2 Dallas Cou…           234625   2453   2552213         53626           91.9
## 3 Tarrant Co…           195518   1798   1983675         62532           98.6
## 4 Bexar Coun…           152231   2040   1892004         53999           80.5
## 5 El Paso Co…           107552   1940    834825         43244          129.
## 6 Collin Cou…            64721    483    914075         90124           70.8
## # … with 2 more variables: deaths_per_1000 <dbl>, death_per_case <dbl>
```

```
datatable(cases_TX_select) %>% formatRound(6:7, 4) %>% formatPercentage(8, 2)
```

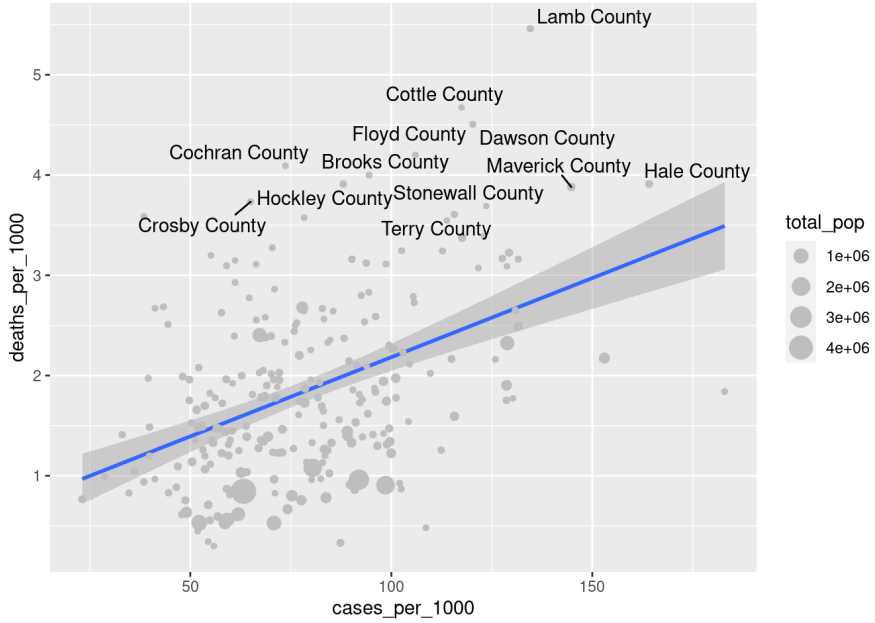
Show 10 entries Search:

| | county_name | confirmed_cases | deaths | total_pop | median_income | cases_per_1000 | deaths_per_1000 | death_pe |
|---|---|---|---|---|---|---|---|---|
| 1 | Harris County | 286356 | 3825 | 4525519 | 57791 | 63.2758 | 0.8452 | |
| 2 | Dallas County | 234625 | 2453 | 2552213 | 53626 | 91.9300 | 0.9611 | |
| 3 | Tarrant County | 195518 | 1798 | 1983675 | 62532 | 98.5635 | 0.9064 | |
| 4 | Bexar County | 152231 | 2040 | 1892004 | 53999 | 80.4602 | 1.0782 | |
| 5 | El Paso County | 107552 | 1940 | 834825 | 43244 | 128.8318 | 2.3238 | |
| 6 | Collin County | 64721 | 483 | 914075 | 90124 | 70.8049 | 0.5284 | |
| 7 | Travis County | 61468 | 626 | 1176584 | 68350 | 52.2428 | 0.5320 | |
| 8 | Hidalgo County | 56455 | 2018 | 839539 | 37097 | 67.2452 | 2.4037 | |
| 9 | Denton County | 46272 | 439 | 781321 | 80290 | 59.2228 | 0.5619 | |
| 10 | Lubbock County | 45600 | 648 | 298042 | 49078 | 152.9986 | 2.1742 | |

Showing 1 to 10 of 240 entries Previous 1 2 3 4 5 … 24 Next

```
ggplot(cases_TX_select, mapping = aes(x = cases_per_1000, y = deaths_per_1000, label = county_name)) +
  geom_smooth(method = lm) +
  geom_point(mapping = aes(size = total_pop), color = "grey") +
  geom_text_repel(data = subset(cases_TX_select, deaths_per_1000 > quantile(deaths_per_1000, .95)))
```

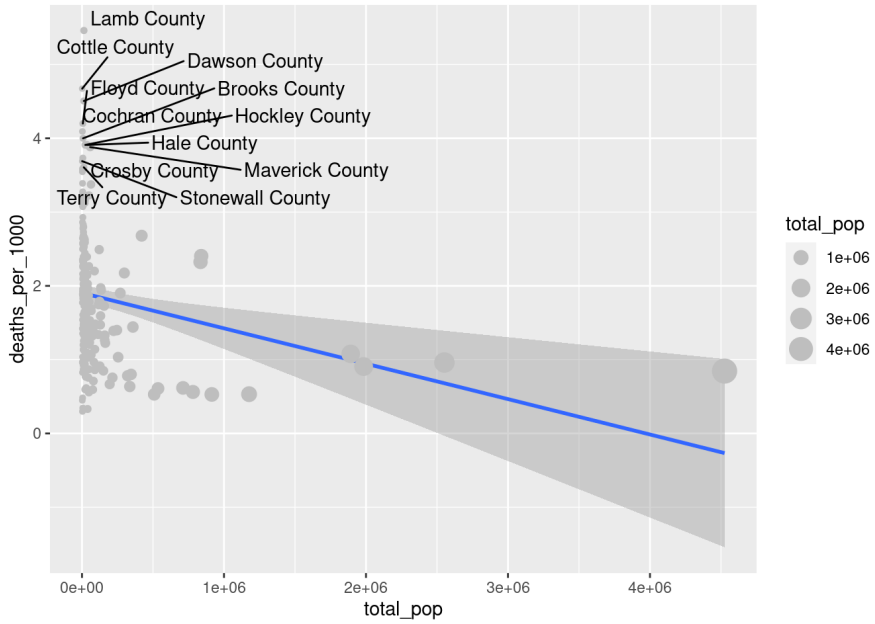
```
## `geom_smooth()` using formula 'y ~ x'
```




## Does death per case depend on population?

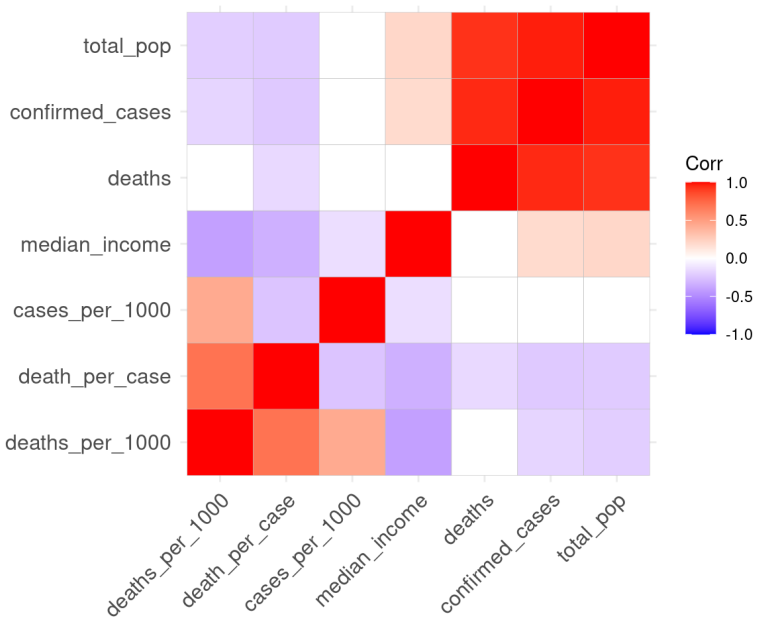
```
ggplot(cases_TX_select, mapping = aes(x= total_pop, y = deaths_per_1000, label = county_name)) +
  geom_smooth(method = lm) +
  geom_point(mapping = aes(size = total_pop), color = "grey") +
  geom_text_repel(data = subset(cases_TX_select, deaths_per_1000 > quantile(deaths_per_1000, .95)))
```

```
## `geom_smooth()` using formula 'y ~ x'
```




## What variables are correlated?

```
cor_TX <- cor(cases_TX_select[,-1])
ggcorrplot(cor_TX, p.mat = cor_pmat(cases_TX_select[,-1]), insig = "blank", hc.order = TRUE)
```




## Plot as a map

See: https://eriqande.github.io/rep-res-web/lectures/making-maps-with-R.html (https://eriqande.github.io/rep-res-web/lectures/making-maps-with-R.html)

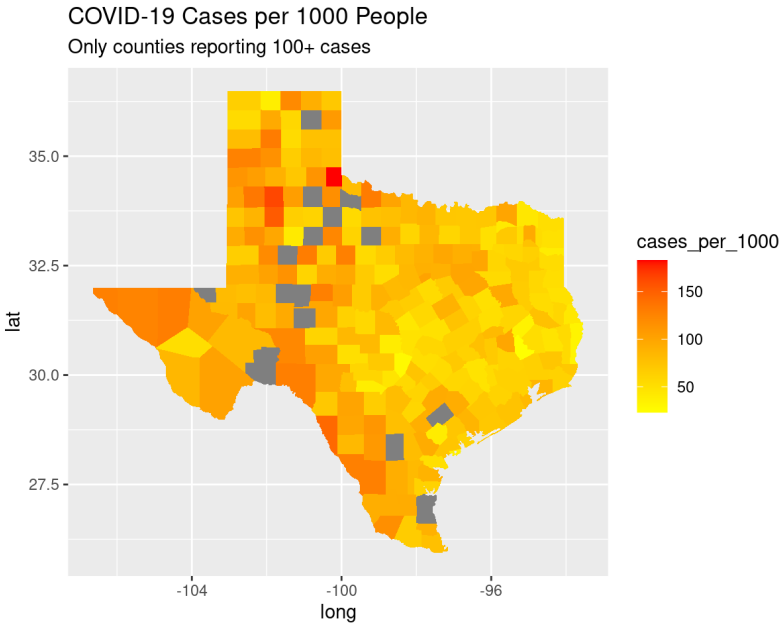
```
counties <- as_tibble(map_data("county"))
counties_TX <- counties %>% dplyr::filter(region == "texas") %>% rename(c(county = subregion))

cases_TX <- cases_TX_select %>% mutate(county = county_name %>% str_to_lower() %>%
    str_replace('\\s+county\\s*$', ''))
```

add variables to map data

```
counties_TX <- counties_TX %>% left_join(cases_TX %>%
    select(c(county, cases_per_1000, deaths_per_1000, death_per_case)))
```

```
## Joining, by = "county"
```

```
ggplot(counties_TX, aes(long, lat, label = county)) +
  geom_polygon(aes(group = group, fill = cases_per_1000)) +
 # geom_text_repel(data = counties_TX %>% filter(complete.cases(.)) %>% group_by(county) %>%
#    summarize(long = mean(long), lat = mean(lat)) %>% mutate(county = str_to_title(county))) +
  coord_quickmap() +
  scale_fill_gradient(low="yellow", high="red") +
  labs(title = "COVID-19 Cases per 1000 People", subtitle = "Only counties reporting 100+ cases")
```




# Look at Dallas County over time. Are we flattening the curve?

Source: https://console.cloud.google.com/marketplace/details/usafacts-public-data/covid19-us-cases?filter=solution-type:dataset&filter=category:covid19&id=3eaff9c5-fbaf-47bb-a441-89db1e1395ab (https://console.cloud.google.com/marketplace/details/usafacts-public-data/covid19-us-cases?filter=solution-type:dataset&filter=category:covid19&id=3eaff9c5-fbaf-47bb-a441-89db1e1395ab)

```
SELECT *
FROM `bigquery-public-data.covid19_usafacts.summary` covid19 WHERE state = "TX"
```

```
cases_TX <- read_csv("COVID-19_cases_TX.csv")
```

```
## Parsed with column specification:
## cols(
##   county_fips_code = col_character(),
##   county_name = col_character(),
##   state = col_character(),
##   state_fips_code = col_double(),
##   date = col_date(format = ""),
##   confirmed_cases = col_double(),
##   deaths = col_double()
## )
```

```
cases_TX
```

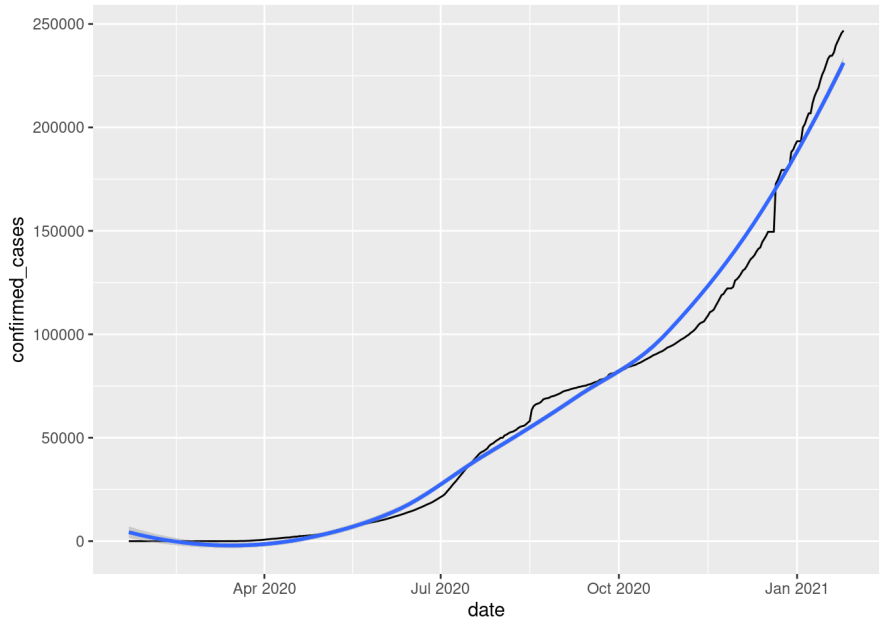
```
## # A tibble: 94,350 x 7
##    county_fips_code county_name state state_fips_code date       confirmed_cases
##    <chr>            <chr>       <chr>           <dbl> <date>               <dbl>
##  1 00000            Statewide … TX                 48 2020-01-22               0
##  2 00000            Statewide … TX                 48 2020-01-23               0
##  3 00000            Statewide … TX                 48 2020-01-24               0
##  4 00000            Statewide … TX                 48 2020-01-25               0
##  5 00000            Statewide … TX                 48 2020-01-26               0
##  6 00000            Statewide … TX                 48 2020-01-27               0
##  7 00000            Statewide … TX                 48 2020-01-28               0
##  8 00000            Statewide … TX                 48 2020-01-29               0
##  9 00000            Statewide … TX                 48 2020-01-30               0
## 10 00000            Statewide … TX                 48 2020-01-31               0
## # … with 94,340 more rows, and 1 more variable: deaths <dbl>
```

```
cases_Dallas <- cases_TX %>% filter(county_name == "Dallas County" & state == "TX")
dim(cases_Dallas)
```

```
## [1] 370   7
```

```
ggplot(cases_Dallas, aes(x = date, y = confirmed_cases)) + geom_line() + geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




You probably should look at the new cases per day # The Effect of Staying at Home

Source: https://www.google.com/covid19/mobility/index.html (https://www.google.com/covid19/mobility/index.html)

```
mobility <- read_csv("Global_Mobility_Report.csv")
```

```
## Parsed with column specification:
## cols(
##   country_region_code = col_character(),
##   country_region = col_character(),
##   sub_region_1 = col_character(),
##   sub_region_2 = col_logical(),
##   metro_area = col_logical(),
##   iso_3166_2_code = col_character(),
##   census_fips_code = col_logical(),
##   date = col_date(format = ""),
##   retail_and_recreation_percent_change_from_baseline = col_double(),
##   grocery_and_pharmacy_percent_change_from_baseline = col_double(),
##   parks_percent_change_from_baseline = col_double(),
##   transit_stations_percent_change_from_baseline = col_double(),
##   workplaces_percent_change_from_baseline = col_double(),
##   residential_percent_change_from_baseline = col_double()
## )
```

```
## Warning: 4199216 parsing failures.
##  row        col           expected                  actual                          file
## 3036 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3037 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3038 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3039 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3040 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## .... .......... .................. ...................... ..........................
## See problems(...) for more details.
```

```
mobility <- read_csv("Global_Mobility_Report.csv", col_types =  cols(sub_region_2 = col_character()))
```

```
## Warning: 874357 parsing failures.
##  row        col           expected                  actual                          file
## 3036 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3037 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3038 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3039 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## 3040 metro_area 1/0/T/F/TRUE/FALSE Kabul Metropolitan Area 'Global_Mobility_Report.csv'
## .... .......... .................. ...................... ..........................
## See problems(...) for more details.
```

```
mobility <- mobility %>% mutate_if(is.character, factor)
dim(mobility)
```

```
## [1] 3991405      14
```

```
head(mobility)
```

```
## # A tibble: 6 x 14
##   country_region_… country_region sub_region_1 sub_region_2 metro_area
##   <fct>            <fct>          <fct>        <fct>        <lgl>
## 1 AE               United Arab E… <NA>         <NA>         NA
## 2 AE               United Arab E… <NA>         <NA>         NA
## 3 AE               United Arab E… <NA>         <NA>         NA
## 4 AE               United Arab E… <NA>         <NA>         NA
## 5 AE               United Arab E… <NA>         <NA>         NA
## 6 AE               United Arab E… <NA>         <NA>         NA
## # … with 9 more variables: iso_3166_2_code <fct>, census_fips_code <lgl>,
## #   date <date>, retail_and_recreation_percent_change_from_baseline <dbl>,
## #   grocery_and_pharmacy_percent_change_from_baseline <dbl>,
## #   parks_percent_change_from_baseline <dbl>,
## #   transit_stations_percent_change_from_baseline <dbl>,
## #   workplaces_percent_change_from_baseline <dbl>,
## #   residential_percent_change_from_baseline <dbl>
```

```
summary(mobility)
```

```
##  country_region_code        country_region
##  US     : 869690     United States : 869690
##  BR     : 638745     Brazil        : 638745
##  IN     : 227589     India         : 227589
##  TR     : 180337     Turkey        : 180337
##  GB     : 142583     United Kingdom: 142583
##  (Other):1929917     Argentina     : 140061
##  NA's   :   2544     (Other)       :1792400
##                      sub_region_1               sub_region_2
##  State of São Paulo        : 119671   Washington County:   9110
##  State of Minas Gerais     :  82551   Jefferson County :   7555
##  Texas                     :  64162   Franklin County  :   7122
##  State of Rio Grande do Sul:  49369   Jackson County   :   6516
##  State of Paraná           :  48479   Lincoln County   :   6213
##  (Other)                   :3558930   (Other)          :3288343
##  NA's                      :  68243   NA's             : 666546
##  metro_area     iso_3166_2_code   census_fips_code      date
##  Mode:logical   AE-AJ  :    343   Mode:logical     Min.   :2020-02-15
##  NA's:3991405   AE-AZ  :    343   NA's:3991405     1st Qu.:2020-05-12
##                 AE-DU  :    343                    Median :2020-08-03
##                 AE-FU  :    343                    Mean   :2020-08-06
##                 AE-RK  :    343                    3rd Qu.:2020-11-03
##                 (Other): 714316                    Max.   :2021-01-22
##                 NA's   :3275374
##  retail_and_recreation_percent_change_from_baseline
##  Min.   :-100.0
##  1st Qu.: -41.0
##  Median : -19.0
##  Mean   : -23.2
##  3rd Qu.:  -4.0
##  Max.   : 545.0
##  NA's   :1478424
##  grocery_and_pharmacy_percent_change_from_baseline
##  Min.   :-100
##  1st Qu.: -14
##  Median :  -2
##  Mean   :  -3
##  3rd Qu.:   9
##  Max.   : 615
##  NA's   :1564666
##  parks_percent_change_from_baseline
##  Min.   :-100.0
##  1st Qu.: -44.0
##  Median : -17.0
##  Mean   :  -9.5
##  3rd Qu.:  11.0
##  Max.   :1206.0
##  NA's   :2080860
##  transit_stations_percent_change_from_baseline
##  Min.   :-100.0
##  1st Qu.: -48.0
##  Median : -28.0
##  Mean   : -27.2
##  3rd Qu.:  -7.0
##  Max.   : 554.0
##  NA's   :1973496
##  workplaces_percent_change_from_baseline
##  Min.   :-100.00
##  1st Qu.: -32.00
##  Median : -19.00
##  Mean   : -20.07
##  3rd Qu.:  -5.00
##  Max.   : 260.00
##  NA's   :189760
##  residential_percent_change_from_baseline
##  Min.   :-46.0
##  1st Qu.:  4.0
##  Median :  8.0
##  Mean   :  9.4
##  3rd Qu.: 14.0
##  Max.   : 65.0
##  NA's   :1678955
```

```
mobility_Dallas <- mobility %>% filter(sub_region_1 == "Texas" & sub_region_2 == "Dallas County")
dim(mobility_Dallas)
```

```
## [1] 343  14
```

```
mobility_Dallas
```

```
## # A tibble: 343 x 14
##    country_region_… country_region sub_region_1 sub_region_2 metro_area
##    <fct>            <fct>          <fct>        <fct>        <lgl>
##  1 US               United States  Texas        Dallas Coun… NA
##  2 US               United States  Texas        Dallas Coun… NA
##  3 US               United States  Texas        Dallas Coun… NA
##  4 US               United States  Texas        Dallas Coun… NA
##  5 US               United States  Texas        Dallas Coun… NA
##  6 US               United States  Texas        Dallas Coun… NA
##  7 US               United States  Texas        Dallas Coun… NA
##  8 US               United States  Texas        Dallas Coun… NA
##  9 US               United States  Texas        Dallas Coun… NA
## 10 US               United States  Texas        Dallas Coun… NA
## # … with 333 more rows, and 9 more variables: iso_3166_2_code <fct>,
## #   census_fips_code <lgl>, date <date>,
## #   retail_and_recreation_percent_change_from_baseline <dbl>,
## #   grocery_and_pharmacy_percent_change_from_baseline <dbl>,
## #   parks_percent_change_from_baseline <dbl>,
## #   transit_stations_percent_change_from_baseline <dbl>,
## #   workplaces_percent_change_from_baseline <dbl>,
## #   residential_percent_change_from_baseline <dbl>
```

```
ggplot(mobility_Dallas, mapping = aes(x = date, y = retail_and_recreation_percent_change_from_baseline)) + geom_li
ne() + geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```