Covid

Covid data analysis DK overview

```
dc
```

```
## # A tibble: 211 x 98
##    date_sample Roskilde Copenhagen Aarhus Frederiksberg Kalundborg Middelfart
##    <date>         <dbl>      <dbl>  <dbl>         <dbl>      <dbl>      <dbl>
##  1 2020-02-26         1          0      0             0          0          0
##  2 2020-02-27         0          1      0             0          0          0
##  3 2020-02-28         0          0      1             0          0          0
##  4 2020-03-01         0          1      0             0          0          0
##  5 2020-03-02         0          1      0             1          1          1
##  6 2020-03-03         0          4      0             0          0          0
##  7 2020-03-04         0          1      0             0          0          0
##  8 2020-03-05         0          2      0             2          0          0
##  9 2020-03-06         0          0      0             0          0          0
## 10 2020-03-07         0          1      1             1          0          0
## # ... with 201 more rows, and 91 more variables: Gentofte <dbl>, Solrød <dbl>,
## #   Allerød <dbl>, Dragør <dbl>, Helsingør <dbl>, Hillerød <dbl>,
## #   Rudersdal <dbl>, Næstved <dbl>, Aalborg <dbl>, Vallensbæk <dbl>,
## #   Esbjerg <dbl>, Favrskov <dbl>, Frederikssund <dbl>, Furesø <dbl>,
## #   Greve <dbl>, Holbæk <dbl>, Kolding <dbl>, Odense <dbl>, Silkeborg <dbl>,
## #   Skanderborg <dbl>, Vejle <dbl>, Aabenraa <dbl>, Ballerup <dbl>,
## #   Egedal <dbl>, Faxe <dbl>, Fredericia <dbl>, Faaborg-Midtfyn <dbl>, ...
```

```
dt
```

```
## # A tibble: 230 x 101
##    PrDate_adjusted Copenhagen Frederiksberg Dragør Tårnby Albertslund Ballerup
##    <date>               <dbl>         <dbl>  <dbl>  <dbl>       <dbl>    <dbl>
##  1 2020-01-27               0             0      0      0           0        0
##  2 2020-01-29               1             0      0      0           0        0
##  3 2020-02-02               0             0      0      0           0        0
##  4 2020-02-03               0             0      0      0           0        0
##  5 2020-02-05               0             0      0      0           0        0
##  6 2020-02-06               0             0      0      0           0        0
##  7 2020-02-07               0             0      0      0           0        0
##  8 2020-02-12               0             0      0      0           0        0
##  9 2020-02-13               0             0      0      0           0        0
## 10 2020-02-14               0             0      0      0           0        0
## # ... with 220 more rows, and 94 more variables: Brøndby <dbl>, Gentofte <dbl>,
## #   Gladsaxe <dbl>, Glostrup <dbl>, Herlev <dbl>, Hvidovre <dbl>,
## #   Høje-Taastrup <dbl>, Ishøj <dbl>, Lyngby-Taarbæk <dbl>, Rødovre <dbl>,
## #   Vallensbæk <dbl>, Allerød <dbl>, Egedal <dbl>, Fredensborg <dbl>,
## #   Frederikssund <dbl>, Furesø <dbl>, Gribskov <dbl>, Halsnæs <dbl>,
## #   Helsingør <dbl>, Hillerød <dbl>, Hørsholm <dbl>, Rudersdal <dbl>,
## #   Bornholm <dbl>, Christiansø <dbl>, Greve <dbl>, Køge <dbl>, ...
```

```
dsize
```

```
## # A tibble: 99 x 6
##    `Kommune_(id)` `Kommune_(navn)` Antal_testede `Antal_bekræfte~ Befolkningstal
##             <dbl> <chr>                    <dbl>            <dbl>          <dbl>
##  1            101 København               271.              4.32          632.
##  2            147 Frederiksberg            44.1            739             104.
##  3            151 Ballerup                 21.5            304              48.6
##  4            153 Brøndby                  15.0            344              35.1
##  5            155 Dragør                    5.19            34              14.5
##  6            157 Gentofte                 29.1            401              74.8
##  7            159 Gladsaxe                 27.8            444              69.3
##  8            161 Glostrup                  9.45           251              23.1
##  9            163 Herlev                   12.6            259              29.0
## 10            165 Albertslund              11.4            249              27.7
## # ... with 89 more rows, and 1 more variable:
## #   Kumulativ_incidens_(per_100000) <dbl>
```

```
# make data tidy
dc <- dc %>%
  #use pivot_longer to transpose all columns (apart from date_sample) to two columns ('kommune','testsC
  pivot_longer(cols = !date_sample, names_to = "kommune", values_to = "casesDiagnosed") %>% arrange(komm
dc
```

```
## # A tibble: 20,467 x 3
##    date_sample kommune  casesDiagnosed
##    <date>      <chr>             <dbl>
##  1 2020-02-26  Aabenraa              0
##  2 2020-02-27  Aabenraa              0
##  3 2020-02-28  Aabenraa              0
##  4 2020-03-01  Aabenraa              0
##  5 2020-03-02  Aabenraa              0
##  6 2020-03-03  Aabenraa              0
##  7 2020-03-04  Aabenraa              0
##  8 2020-03-05  Aabenraa              0
##  9 2020-03-06  Aabenraa              0
## 10 2020-03-07  Aabenraa              0
## # ... with 20,457 more rows
```

```
dt <- dt %>%
  #use pivot_longer to transpose all columns (apart from PrDate_adjusted) to two columns ('kommune','te
  pivot_longer(cols = !PrDate_adjusted, names_to = "kommune", values_to = "testsConducted") %>%
  arrange(kommune, PrDate_adjusted) %>%
  rename(date_sample = PrDate_adjusted) %>%
  filter(!(kommune == "X101"))
```

```
# check there are no NA in the sets dc and dt , dc$casesDiagnosed, and dt$testsConducted
sum(is.na(dc$casesDiagnosed))
```

```
## [1] 0
```

```
sum(is.na(dt$testsConducted))
```

```
## [1] 0
```

```
sum(is.na(dsize$Befolkningstal))
```

```
## [1] 0
```

```r
# make data compatible rename columns for merging
dsize <- dsize %>%
  select(contains("Kom"), Befolkningstal) %>%
  rename("kID" = `Kommune_(id)`, "kommune" = `Kommune_(navn)`, population = Befolkningstal) %>%
  mutate(population = population * 1000) %>%
  select(-kID)
```

```r
# check Kommune name can be used as key
dc[!(dc$kommune %in% dsize$kommune), ]$kommune
```

```
##   [1] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##   [6] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [11] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [16] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [21] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [26] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [31] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [36] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [41] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [46] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [51] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [56] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [61] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [66] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [71] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [76] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [81] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [86] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [91] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [96] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [101] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [106] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [111] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [116] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [121] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [126] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [131] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [136] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [141] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [146] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [151] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [156] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [161] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [166] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [171] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [176] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [181] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [186] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [191] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [196] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [201] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [206] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [211] "Copenhagen"
```

```
dsize[!(dsize$kommune %in% dt$kommune), ]$kommune
```

```
## [1] "København"
```

```
dt[!(dt$kommune %in% dsize$kommune), ]$kommune
```

```
##   [1] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##   [6] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [11] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [16] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [21] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [26] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [31] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [36] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [41] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [46] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [51] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [56] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [61] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [66] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [71] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [76] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [81] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [86] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [91] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
##  [96] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [101] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [106] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [111] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [116] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [121] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [126] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [131] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [136] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [141] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [146] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [151] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [156] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [161] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [166] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [171] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [176] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [181] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [186] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [191] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [196] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [201] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [206] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [211] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [216] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [221] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
## [226] "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen" "Copenhagen"
```

```
# ooops!
# for some reason one data frame uses Københaven the other Copenhagen), - use str_replace to change "Kø
```

```
dsize$kommune <- str_replace(dsize$kommune, "København", "Copenhagen")

# check again keys are clean OK re-run rows 35-37

# merge data together
dc <- merge(dc, dsize)
dt <- merge(dt, dsize)
```

```
# create new variables within mutate
# (e.g. rolling aggregates with roll_sum - cases of the last seven days) -
# numbers per 1000 people
# use the lag function to compute the differences between a days (7day sum) value and the one from 7 da
#
dc <- dc %>%
  group_by(kommune) %>%
  arrange(kommune,date_sample) %>%
  mutate(dcr100k = casesDiagnosed / (population / 100000),
         dcr7d = roll_sum(casesDiagnosed, width = 7, min_obs = 1),
         dcr7dPer100k= dcr7d /(population / 100000),
         dcr7dPer100kCh1 = dcr7dPer100k - lag(dcr7dPer100k,1),
         dcr7dPer100kCh3 = dcr7dPer100k - lag(dcr7dPer100k,3),
         dcr7dPer100kCh7 = dcr7dPer100k - lag(dcr7dPer100k,7)
         )

dc
```

```
## # A tibble: 20,467 x 10
## # Groups:   kommune [97]
##    kommune  date_sample casesDiagnosed population dcr100k dcr7d dcr7dPer100k
##    <chr>    <date>               <dbl>      <dbl>   <dbl> <dbl>         <dbl>
##  1 Aabenraa 2020-02-26               0      58761       0     0             0
##  2 Aabenraa 2020-02-27               0      58761       0     0             0
##  3 Aabenraa 2020-02-28               0      58761       0     0             0
##  4 Aabenraa 2020-03-01               0      58761       0     0             0
##  5 Aabenraa 2020-03-02               0      58761       0     0             0
##  6 Aabenraa 2020-03-03               0      58761       0     0             0
##  7 Aabenraa 2020-03-04               0      58761       0     0             0
##  8 Aabenraa 2020-03-05               0      58761       0     0             0
##  9 Aabenraa 2020-03-06               0      58761       0     0             0
## 10 Aabenraa 2020-03-07               0      58761       0     0             0
## # ... with 20,457 more rows, and 3 more variables: dcr7dPer100kCh1 <dbl>,
## #   dcr7dPer100kCh3 <dbl>, dcr7dPer100kCh7 <dbl>
```

```
dt <- dt %>%
  group_by(kommune) %>%
  arrange(kommune,date_sample) %>%
  mutate(tr100k = testsConducted / (population / 100000),
         tr7d = roll_sum(testsConducted, width = 7, min_obs = 1),
         tr7dP100k = tr7d / (population / 100000),
         tr7dP100kCh1 = tr7dP100k - lag(tr7dP100k,1),
         tr7dP100kCh3 = tr7dP100k - lag(tr7dP100k,3),
         tr7dP100kCh7 = tr7dP100k - lag(tr7dP100k,7)
         )
dt
```

```
## # A tibble: 22,770 x 10
## # Groups:   kommune [99]
##    kommune  date_sample testsConducted population tr100k  tr7d tr7dP100k
##    <chr>    <date>               <dbl>      <dbl>  <dbl> <dbl>     <dbl>
##  1 Aabenraa 2020-01-27               0      58761      0     0         0
##  2 Aabenraa 2020-01-29               0      58761      0     0         0
##  3 Aabenraa 2020-02-02               0      58761      0     0         0
##  4 Aabenraa 2020-02-03               0      58761      0     0         0
##  5 Aabenraa 2020-02-05               0      58761      0     0         0
##  6 Aabenraa 2020-02-06               0      58761      0     0         0
##  7 Aabenraa 2020-02-07               0      58761      0     0         0
##  8 Aabenraa 2020-02-12               0      58761      0     0         0
##  9 Aabenraa 2020-02-13               0      58761      0     0         0
## 10 Aabenraa 2020-02-14               0      58761      0     0         0
## # ... with 22,760 more rows, and 3 more variables: tr7dP100kCh1 <dbl>,
## #   tr7dP100kCh3 <dbl>, tr7dP100kCh7 <dbl>
```

```
df <- merge(dc, dt)
```

```
# merge test with case data (on shared keys - date_sample+kommune)


#create test positive variables
df <- df %>%
  mutate(
    PosTestRateOnDay = dcr100k / tr100k,
    PosTestRate7d = dcr7dPer100k / tr7dP100k,
  )


# check there were no diagnosis without tests
sum(df$casesDiagnosed > df$testsConducted)
```

```
## [1] 0
```

Covid data analysis DK overview