hpam7660 Cancer Alley Analytic Data File

Grace Arkfeld

2024-02-22

```r
library(readr)
la_mort <-
  read_csv("https://www.dropbox.com/scl/fi/fzsnhfd3lq80v2o3sag6c/la_mort.csv?rlkey=h1vyjm2b8ppgejgsg3e8
```

```
## Rows: 642696 Columns: 29
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): stocr, strsd, stbrth, brthr, sex, marstat, ucod
## dbl (22): restatus, cntyocr, popcntyocr, cntyrsd, popcntyresd, educ1989, edu...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
la_mort$cancer_parish <- ifelse(la_mort$cntyrsd %in% c(5, 33, 47, 51, 71, 89, 93, 95, 121), 1, 0)
```

```r
table(la_mort$cancer_parish)
```

```
##
##      0      1
## 445138 197558
```

```r
table(la_mort$cntyrsd[la_mort$cancer_parish == 1])
```

```
##
##      5     33     47     51     71     89     93     95    121
## 10217  55300   4761  61822  47752   5963   2946   5844   2953
```

```r
table(la_mort$cancer_parish)
```

```
##
##      0      1
## 445138 197558
```

```r
table(la_mort$cntyrsd[la_mort$cancer_parish == 1])
```

```
##
##      5     33     47     51     71     89     93     95    121
## 10217  55300   4761  61822  47752   5963   2946   5844   2953
```

```r
table(la_mort$cancer39)
```

```
## Warning: Unknown or uninitialised column: 'cancer39'.
```

```
## < table of extent 0 >
```

```r
la_mort$cancer39 <- ifelse(la_mort$ucr39 %in% c(5:15), 1, 0)
```

```r
la_mort$cancer113 <- ifelse(la_mort$ucr113 %in% c(19:44), 1, 0)
```

```r
la_mort$cancer39 <- ifelse(la_mort$ucr39 %in% c(5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15), 1, 0)
```

```r
table(la_mort$cancer39)
```

```
##
##      0      1
## 504019 138677
```

```r
la_mort$cancer39 <- ifelse(la_mort$ucr39 %in% c(5:15), 1, 0)
```

```r
table(la_mort$ucr39[la_mort$cancer113 == 1 & la_mort$cancer39 == 0])
```

```
##
##   37
## 3070
```

```r
table(la_mort$ucod[la_mort$cancer113 == 1 & la_mort$cancer39 == 0])
```

```
##
## D000 D010 D022 D037 D049 D051 D059 D070 D075 D090 D120 D126 D131 D133 D134 D136
##    1    7    8    1    1    2    1    1    2    2    1    6    3    1   13    2
## D137 D143 D144 D150 D151 D164 D169 D177 D179 D180 D181 D219 D239 D259 D269  D27
##    5    1    1   10    7    1    1    1    1   11   10    6    3    1    2    2
## D282 D291 D320 D329 D332 D333 D350 D351 D352 D361 D369 D370 D371 D372 D374 D375
##    2    1   26   97   11    3   12    2   17    1    2   12   17    7   28    8
## D376 D377 D379 D380 D381 D383 D384 D385 D386 D390 D391 D392 D399 D400 D407 D410
##   55   40    4   12  157    8    1    1    4   13   10    1    1   12    1   22
## D411 D412 D414 D429 D430 D431 D432 D433 D434 D439 D440 D441 D443 D444 D447 D448
##    1    1   35    3   36   22  561    1    4    1    4    6   21   15    2    1
## D449  D45 D462 D464 D467 D469 D471 D472 D473 D477 D479 D480 D481 D482 D483 D484
##    4   49    1   17    1 1084  214   13    9    1   14   15   74    1    6    8
## D485 D486 D487 D489
##    1   16   39   86
```

```r
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
```

```r
parish_count <- la_mort %>%
  group_by(cntyrsd, cancer_parish, year) %>%
  summarize(cancer39 = sum(cancer39, na.rm = TRUE))
```

```
## 'summarise()' has grouped output by 'cntyrsd', 'cancer_parish'. You can
## override using the '.groups' argument.
```

```r
summary(parish_count$cancer39)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     3.0    42.0    74.0   144.5   159.0   992.0
```

```r
la_pop <-
  read_csv("https://www.dropbox.com/scl/fi/650k1obpczky6bwa19ex6/la_county_pop.csv?rlkey=0aokd9m76q7mxw
```

```
## Rows: 24320 Columns: 23
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr  (3): stname, ctyname, agegrp
## dbl (20): state, county, year, tot_pop, tot_male, tot_female, wa_male, wa_fe...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
parish_count <- parish_count %>%
  rename(county = cntyrsd)
```

```r
la_joined <- parish_count %>%
  inner_join(la_pop, by = c("county", "year"))
```