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

la_mort_age <- la_mort %>%
  filter(age != 9999)
la_mort_age$age <- ifelse(la_mort_age$age < 2000, la_mort_age$age - 1000, 0)

age_breaks <- c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, Inf)
age_labels <- c("0_4", "5_9", "10_14", "15_19", "20_24", "25_29", "30_34", "35_39",
                "40_44", "45_49", "50_54", "55_59", "60_64", "65_69", "70_74",
                "75_79", "80_84", "85+")
la_mort_age$agegrp <- as.character(cut(la_mort_age$age, breaks = age_breaks, labels = age_labels, right

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.
```

```
parish_count_age <- la_mort_age %>%
  group_by(cntyrsd, cancer_parish, agegrp, year) %>%
  summarize(cancer39 = sum(cancer39, na.rm = TRUE))
```

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

```
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.
```

```
# Check column names in both datasets
colnames(parish_count)
```

```
## [1] "cntyrsd"       "cancer_parish" "year"          "cancer39"
```

```
colnames(la_pop)
```

```
##  [1] "state"      "county"     "stname"     "ctyname"    "year"
##  [6] "agegrp"     "tot_pop"    "tot_male"   "tot_female" "wa_male"
## [11] "wa_female"  "ba_male"    "ba_female"  "ia_male"    "ia_female"
## [16] "aa_male"    "aa_female"  "na_male"    "na_female"  "tom_male"
## [21] "tom_female" "h_male"     "h_female"
```

```r
library(dplyr)

# Correct join operation without renaming columns
la_joined <- parish_count %>%
  inner_join(la_pop, by = c("cancer_parish" = "county", "year"))
```

```
## Warning in inner_join(., la_pop, by = c(cancer_parish = "county", "year")): Detected an unexpected ma
## i Row 31 of 'x' matches multiple rows in 'y'.
## i Row 77 of 'y' matches multiple rows in 'x'.
## i If a many-to-many relationship is expected, set 'relationship =
##   "many-to-many"' to silence this warning.
```

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

stnrd_pop <-
  read_csv("https://www.dropbox.com/scl/fi/xzd2o5lza237so6vamqwb/stnrd_pop.csv?rlkey=zp90au2tuq6eptvi1y
```

```
## Rows: 18 Columns: 2
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr (1): agegrp
## dbl (1): stnrd_pop
##
## 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_joined_stnrd <- la_joined %>%
  inner_join(stnrd_pop, by = "agegrp")

la_joined_stnrd$stnrd_pop_weight <- (la_joined_stnrd$stnrd_pop) / (sum(stnrd_pop$stnrd_pop))

la_joined_stnrd$cancer_rate_adj <- ((la_joined_stnrd$cancer39) / (la_joined_stnrd$tot_pop / 100000)) * 

parish_rates <- la_joined_stnrd %>%
  group_by(cntyrsd, cancer_parish, year) %>%
  summarize(cancer_rate_adj = sum(cancer_rate_adj, na.rm = TRUE), cancer39 = sum(cancer39), tot_pop =
            sum(tot_pop))
```

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

```r
parish_rates$cancer_rate_crude <- (parish_rates$cancer39) / (parish_rates$tot_pop / 100000)

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

```
## Warning in inner_join(., la_pop, by = c(cancer_parish = "county", "year")): Detected an unexpected ma
## i Row 31 of 'x' matches multiple rows in 'y'.
## i Row 77 of 'y' matches multiple rows in 'x'.
## i If a many-to-many relationship is expected, set 'relationship =
##   "many-to-many"' to silence this warning.
```

```r
# Assuming la_joined_all exists and contains the correct data
parish_cancer_2019 <- subset(la_joined_all, year == 2019)

library(knitr)

# Adjust the column names to those that exist in your dataframe
kable(parish_cancer_2019[, c("cancer_parish", "cancer39")])
```

cancer_parish	cancer39
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	65
1	65
1	65
1	65
1	65

cancer_parish	cancer39
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731

cancer_parish	cancer39
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78

cancer__parish	cancer39
1	78
1	78
1	78
1	78
1	78
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41

```r
library(knitr)

# Adjusting the code to use existing columns
kable(parish_cancer_2019[, c("cancer_parish", "cancer39")])
```

cancer__parish	cancer39
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162
1	162

cancer_parish	cancer39
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	741
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	65
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915
1	915

cancer_parish	cancer39
1	915
1	915
1	915
1	915
1	915
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	731
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	99
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39

cancer__parish	cancer39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	39
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	78
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41
1	41

```r
parish_cancer_2019 <- subset(parish_rates, year == 2019)
kable(parish_cancer_2019[, c("cntyrsd", "cancer_rate_crude", "cancer_rate_adj")])
```

cntyrsd	cancer_rate_crude	cancer_rate_adj
1	225.0551	182.5548
3	224.8501	181.6644
5	127.7623	131.1103
7	267.7050	186.3922
9	249.1405	188.7601
11	256.6296	195.5435
13	297.9484	168.9799
15	178.9962	156.0769
17	232.0832	175.2182
19	209.7648	179.7991
21	365.1767	183.9878
23	360.5375	106.0562
25	293.9555	119.8748
27	247.8402	145.5867
29	332.9065	205.5863
31	316.2555	171.0102
33	167.8479	149.8723
35	541.2157	327.0177
37	343.6641	202.1083
39	250.9007	188.6733
41	318.8496	182.7571
43	244.4855	164.9050
45	239.6932	189.9145
47	242.3203	157.6160
49	411.5865	227.1434
51	210.7829	159.7797
53	282.8387	193.3757
55	167.0765	154.7989
57	235.3096	173.2434
59	311.3268	170.4757
61	183.6131	157.6739
63	157.5293	145.9635
65	401.3761	226.4382
67	300.4252	192.7607
69	267.3315	187.3432
71	186.8743	157.3156
73	224.9008	188.9408
75	195.4104	144.0914
77	292.6880	168.8143
79	201.8490	160.3180
81	443.4590	204.9235
83	212.5399	155.2236
85	321.3429	192.7206
87	137.5550	122.1881
89	199.9838	176.4014
91	298.3294	162.7734
93	197.8892	138.3994
95	211.0904	161.9156
97	279.9795	209.0688
99	200.8799	157.1110
101	251.5519	191.4809
103	206.2739	161.6818

cntyrsd	cancer_rate_crude	cancer_rate_adj
105	204.6794	183.3851
107	575.6857	190.4850
109	246.3126	193.8026
111	377.7371	187.6486
113	244.1977	153.0439
115	228.4710	225.5663
117	275.1419	207.0287
119	250.5023	171.7850
121	178.1911	138.3193
123	270.4108	159.7352
125	232.8080	128.8255
127	476.4318	234.1971

```r
parish_rates$pop_weight <- (parish_rates$cancer_rate_adj) * (parish_rates$tot_pop)
cancer_alley_rates <- parish_rates %>%
  group_by(cancer_parish, year) %>%
  summarize(cancer_rate_adj_wt = sum(pop_weight) / sum(tot_pop))
```

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

```r
kable(cancer_alley_rates)
```

cancer_parish	year	cancer_rate_adj_wt
0	2005	215.9012
0	2006	211.1969
0	2007	199.2163
0	2008	210.5785
0	2009	202.7788
0	2010	198.5223
0	2011	194.5824
0	2012	194.9155
0	2013	191.4183
0	2014	188.3508
0	2015	186.8605
0	2016	178.2077
0	2017	181.0797
0	2018	176.0163
0	2019	174.1137
1	2005	197.2898
1	2006	198.7948
1	2007	199.3910
1	2008	196.7380
1	2009	190.6874
1	2010	191.1738
1	2011	189.7244
1	2012	180.9129
1	2013	181.2483
1	2014	181.1850

cancer_parish	year	cancer_rate_adj_wt
1	2015	166.3009
1	2016	157.8499
1	2017	161.2732
1	2018	153.9050
1	2019	153.9429

```r
cancer_alley <-
  subset(cancer_alley_rates, cancer_parish == 1, select = c(cancer_rate_adj_wt, year)) %>%
  rename(cancer_alley_rate = cancer_rate_adj_wt)
no_cancer_alley <-
  subset(cancer_alley_rates, cancer_parish == 0, select = c(cancer_rate_adj_wt, year)) %>%
  rename(no_cancer_alley_rate = cancer_rate_adj_wt)
cancer_alley_table <- cancer_alley %>%
  inner_join(no_cancer_alley, by = "year")
cancer_alley_table <- cancer_alley_table[,c("year", "cancer_alley_rate", "no_cancer_alley_rate")]
kable(cancer_alley_table)
```

year	cancer_alley_rate	no_cancer_alley_rate
2005	197.2898	215.9012
2006	198.7948	211.1969
2007	199.3910	199.2163
2008	196.7380	210.5785
2009	190.6874	202.7788
2010	191.1738	198.5223
2011	189.7244	194.5824
2012	180.9129	194.9155
2013	181.2483	191.4183
2014	181.1850	188.3508
2015	166.3009	186.8605
2016	157.8499	178.2077
2017	161.2732	181.0797
2018	153.9050	176.0163
2019	153.9429	174.1137