Summarizing Data Cancer Alley

Rebecca Letsinger

Feb 27, 2024

```
library(readr)
library(knitr)
```

```
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.
```

```
show_col_types = FALSE
```

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

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

```
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.
```

```
library(readr)
```

```
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.
```

```
library(dplyr)
```

```
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.
```

```
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)) * 1
```

```
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.
```

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

```
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.
```

```
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.
```

```
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

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

```
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