

# HomeWork

```
baby_data <- make_babynames_dist()
baby_data
```

```
## # A tibble: 1,639,722 x 9
##   year sex   name     n   prop alive_prob count_thousands age_today
##   <dbl> <chr> <chr> <int> <dbl>   <dbl>         <dbl>      <dbl>
## 1  1900 F    Mary  16706 0.0526     0         16.7        114
## 2  1900 F    Helen  6343 0.0200     0         6.34        114
## 3  1900 F    Anna   6114 0.0192     0         6.11        114
## 4  1900 F   Marg~  5304 0.0167     0         5.30        114
## 5  1900 F    Ruth  4765 0.0150     0         4.76        114
## 6  1900 F   Eliz~  4096 0.0129     0         4.10        114
## 7  1900 F   Flor~  3920 0.0123     0         3.92        114
## 8  1900 F   Ethel  3896 0.0123     0         3.90        114
## 9  1900 F   Marie  3856 0.0121     0         3.86        114
## 10 1900 F   Lill~  3414 0.0107     0         3.41        114
## # ... with 1,639,712 more rows, and 1 more variable: est_alive_today <dbl>
```

```
Josephs <- filter(baby_data, name == "Joseph" & sex == "M")
arrange(baby_data, desc(year))
```

```
## # A tibble: 1,639,722 x 9
##   year sex   name     n   prop alive_prob count_thousands age_today
##   <dbl> <chr> <chr> <int> <dbl>   <dbl>         <dbl>      <dbl>
## 1  2010 F   Isab~  22905 0.0117   0.994         22.9         4
## 2  2010 F   Soph~  20639 0.0105   0.994         20.6         4
## 3  2010 F   Emma  17338 0.00885  0.994         17.3         4
## 4  2010 F   Oliv~  17022 0.00869  0.994         17.0         4
## 5  2010 F    Ava   15429 0.00788  0.994         15.4         4
## 6  2010 F   Emily  14268 0.00729  0.994         14.3         4
## 7  2010 F   Abig~  14243 0.00727  0.994         14.2         4
## 8  2010 F   Madi~  13176 0.00673  0.994         13.2         4
## 9  2010 F   Chloe  11750 0.00600  0.994         11.8         4
## 10 2010 F    Mia   10637 0.00543  0.994         10.6         4
## # ... with 1,639,712 more rows, and 1 more variable: est_alive_today <dbl>
```

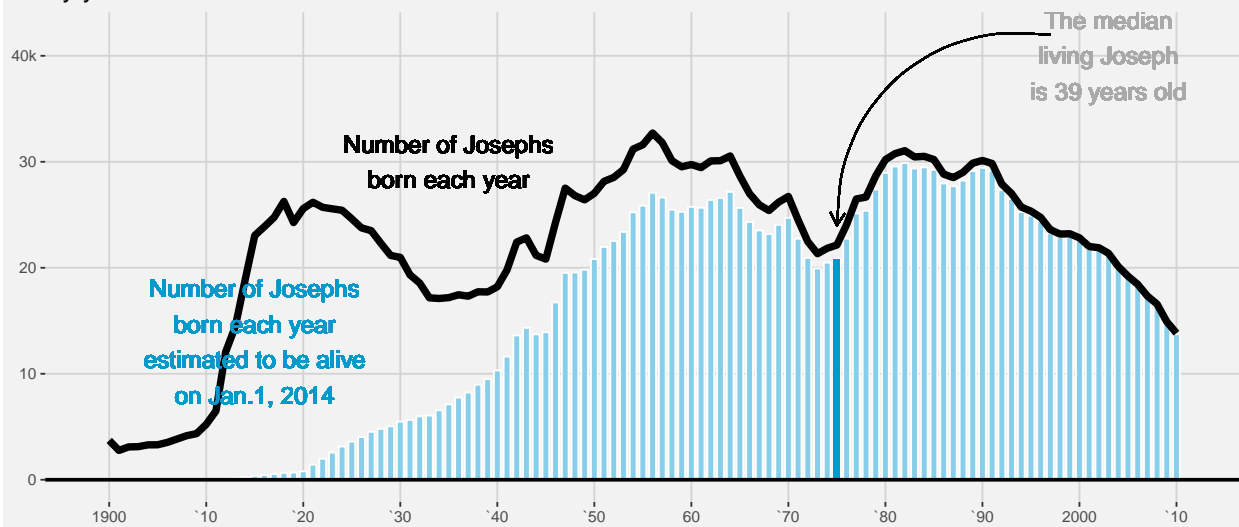
```
median <- wtd.quantile(Josephs$year, Josephs$est_alive_today, probs = 0.5)
```

```
ggplot(data = Josephs, aes(x = year)) +
  geom_bar(aes(y = alive_prob * count_thousands), fill = "skyblue", color = "white", stat = "identity",
  geom_line(aes(y = count_thousands), size = 2) +
  xlab(NULL) +
  ylab(NULL) +
  geom_bar(aes(y = ifelse(year == median, est_alive_today / 1000, 0)), fill = "#0099CC", stat = "identity",
  ggtitle(label = "Age Distribution of American Boys Named Joseph", subtitle = "By year of birth") +
  geom_text(x = 1915, y = 13, label = "Number of Josephs\nborn each year\nestimated to be alive\non Jan",
  geom_text(x = 1935, y = 30, label = "Number of Josephs\nborn each year", size = 5) +
  geom_text(x = 2003, y = 40, label = "The median\nliving Joseph\nis 39 years old", color = "darkgray",
  geom_curve(x = 1997, xend = 1975, y = 42, yend = 24, arrow = arrow(length = unit(0.3,"cm")), curvature = 0.5,
  scale_y_continuous(limits = c(0,42), labels = c("0","10","20","30","40k")) +
  scale_x_continuous(limits = c(1899,2011), breaks = seq(1900,2010,10), labels = c("1900","`10","`20","`30","`40k")) +
  geom_abline(intercept = 0, slope = 0, size = 1) +
```

```
theme(panel.background = element_rect(fill = "grey95"),
      plot.background = element_rect(fill = "grey95"),
      panel.grid.minor.y = element_blank(),
      panel.grid.major.y = element_line(color = "lightgray"),
      panel.grid.major.x = element_line(color = "lightgray"),
      panel.grid.minor.x = element_blank(),
      plot.title = element_text(size = 30),
      plot.subtitle = element_text(size = 15))
```

## Age Distribution of American Boys Named Joseph

By year of birth



# Optional

```
com_m <- baby_data %>%
  filter(sex == "M") %>%
  group_by(name) %>%
  dplyr :: summarize(N = n(), est_sum_alive = sum(est_alive_today),
                    q1_age = 2014 - tryCatch(wtd.quantile(year, est_alive_today, probs = 0.75),
                                             error = function(e) 0),
                    median_age = 2014 - tryCatch(wtd.quantile(year, est_alive_today, probs = 0.5),
                                                  error = function(e) 0),
                    q3_age = 2014 - tryCatch(wtd.quantile(year, est_alive_today, probs = 0.25),
                                             error = function(e) 0)) %>%

  arrange(desc(est_sum_alive)) %>%
  head(25) %>%
  arrange(median_age)

ggplot(data = com_m, aes(x = reorder(name, -median_age), y = median_age)) +
  xlab(NULL) + ylab(NULL) +
  ggtitle(label = "Median Ages for Males With the 25 Most\nCommon Names",
         subtitle = "Among Americans estimated to be alive as of Jan.1, 2014") +
  geom_linerange(aes(ymin = q1_age, ymax = q3_age), color = "skyblue", size = 8, alpha = 0.8) +
  geom_point(fill = "red", color = "white", size = 4, shape = 21) +
  geom_point(aes(y = 59, x = 24), fill = "red", color = "white", size = 4, shape = 21) +
  geom_text(aes(y = 61, x = 24), label = "median") +
  geom_text(aes(y = 27, x = 16), label = "25th") +
  geom_text(aes(y = 52, x = 16), label = "75th percentile") +
```

```
geom_point(aes(y = 25, x = 16), shape = 17, size = 3) +
geom_point(aes(y = 56, x = 16), shape = 17, size = 3) +
theme(panel.background = element_rect(fill = "grey95"),
      plot.background = element_rect(fill = "grey95"),
      plot.title = element_text(size = 30),
      plot.subtitle = element_text(size = 15)) +
coord_flip()
```

