DSCI 401 - HW8

Isabel Heard

12/12/2023

Contents

1

Answer all these questions using the babynames data set from the babynames package in R.

```
library(babynames)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.0     v readr     2.1.2
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.1     v tibble    3.1.8
## v lubridate 1.8.0     v tidyr     1.2.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
```

A

Find all the baby names that are exactly 4 letters long, begin with two vowels (a,e,i,o,u), and end in the letter y. How many are there?

```r
filtered_names <- babynames %>%
  filter(
    nchar(name) == 4,
    substr(name, nchar(name), nchar(name)) == "y",
    grepl("^[aeiou]{2}", name, ignore.case = TRUE))
#print(filtered_names$name)

#Count unique names
number_of_names <- filtered_names %>%
  summarize(count = n_distinct(name))
print(paste("Number of unique names that are exactly 4 letters long, begin with two vowels, and end with
```

```
## [1] "Number of unique names that are exactly 4 letters long, begin with two vowels, and end with the
```
```r
#Print unique names
distinct_names3 <- distinct(filtered_names, name)
print(distinct_names3$name)
```
```
##  [1] "Audy" "Eary" "Eury" "Aimy" "Eimy" "Aury" "Aivy" "Eily" "Aily" "Oumy"
## [11] "Eidy" "Eudy" "Aidy" "Iory" "Aary"
```

B

Find all the babynames that begin and end with the letter g. How many are there?

```r
g_names <- babynames$name[grep("^g.*g$", babynames$name, ignore.case = TRUE)]
number_of_names <- length(g_names)

#Count unique names
unique_count <- length(unique(g_names))
print(paste("Number of unique names that start and end with g:", unique_count))
```
```
## [1] "Number of unique names that start and end with g: 17"
```
```r
#Print unique names
distinct_names <- unique(g_names)
print(distinct_names)
```
```
##  [1] "Gregg"    "Greg"     "Georg"    "Greig"    "Gehrig"   "Graig"
##  [7] "Grayling" "Greyling" "Gig"      "Grieg"    "Gregorg"  "Ging"
## [13] "Giang"    "Gevorg"   "Gaozong"  "Gaosheng" "Garang"
```