Analysis of Toronto Homeless Shelter System*
My subtitle if needed

Huda Sahaf

27 January 2023

First sentence. Second sentence. Third sentence. Fourth sentence.

```r
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.1
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```r
library(janitor)
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```r
library(dplyr)
library(ggplot2)
library(kableExtra)
```

```
Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows
```

```r
library(opendatatoronto)
```

```
# A tibble: 6 x 18
  `_id` date(m~1 popul~2 retur~3 retur~4 newly~5 moved~6 becam~7 activ~8 ageun~9
  <dbl> <chr>    <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1     1 Jan-18   All Po~      46     494    1106     508     862    7958    1233
2     2 Jan-18   Chronic      11      29     317     111     111    2532     223
3     3 Jan-18   Refuge~       4      32     651     328      75    2408     914
4     4 Jan-18   Famili~       0      14     561     321      57    2277    1232
5     5 Jan-18   Youth        12      64     116      44     116     924       0
6     6 Jan-18   Single~      34     416     429     143     689    4757       0
# ... with 8 more variables: `age16-24` <dbl>, `age25-44` <dbl>,
#   `age45-64` <dbl>, age65over <dbl>, gender_male <dbl>, gender_female <dbl>,
#   `gender_transgender,non-binary_or_two_spirit` <dbl>,
#   population_group_percentage <chr>, and abbreviated variable names
#   1: `date(mmm-yy)`, 2: population_group, 3: returned_from_housing,
#   4: returned_to_shelter, 5: newly_identified, 6: moved_to_housing,
#   7: became_inactive, 8: actively_homeless, 9: ageunder16

# A tibble: 6 x 18
  `_id` date(m~1 popul~2 retur~3 retur~4 newly~5 moved~6 becam~7 activ~8 ageun~9
  <dbl> <chr>    <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1   439 Dec-22   Refuge~      31      24     605     263      82    3530     803
2   440 Dec-22   Famili~      27      20     554     263      70    2529    1251
3   441 Dec-22   Youth        12      32     123      29      64     936       0
4   442 Dec-22   Single~      54     448     549     211     429    6943       0
5   443 Dec-22   Non-re~      62     476     621     240     481    6878     470
6   444 Dec-22   Indige~       8     124      52      17      98     978       3
# ... with 8 more variables: `age16-24` <dbl>, `age25-44` <dbl>,
#   `age45-64` <dbl>, age65over <dbl>, gender_male <dbl>, gender_female <dbl>,
#   `gender_transgender,non-binary_or_two_spirit` <dbl>,
#   population_group_percentage <chr>, and abbreviated variable names
#   1: `date(mmm-yy)`, 2: population_group, 3: returned_from_housing,
#   4: returned_to_shelter, 5: newly_identified, 6: moved_to_housing,
#   7: became_inactive, 8: actively_homeless, 9: ageunder16
```

```
# A tibble: 6 x 18
     id date_m~1 popul~2 retur~3 retur~4 newly~5 moved~6 becam~7 activ~8 ageun~9
  <dbl> <chr>    <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1     1 Jan-18   All Po~      46     494    1106     508     862    7958    1233
2     2 Jan-18   Chronic      11      29     317     111     111    2532     223
3     3 Jan-18   Refuge~       4      32     651     328      75    2408     914
4     4 Jan-18   Famili~       0      14     561     321      57    2277    1232
5     5 Jan-18   Youth        12      64     116      44     116     924       0
6     6 Jan-18   Single~      34     416     429     143     689    4757       0
# ... with 8 more variables: age16_24 <dbl>, age25_44 <dbl>, age45_64 <dbl>,
#   age65over <dbl>, gender_male <dbl>, gender_female <dbl>,
#   gender_transgender_non_binary_or_two_spirit <dbl>,
#   population_group_percentage <chr>, and abbreviated variable names
#   1: date_mmm_yy, 2: population_group, 3: returned_from_housing,
#   4: returned_to_shelter, 5: newly_identified, 6: moved_to_housing,
#   7: became_inactive, 8: actively_homeless, 9: ageunder16
```

```r
#### Renaming the columns ####
names(cleaned_shelter_data)
```

```
 [1] "id"
 [2] "date_mmm_yy"
 [3] "population_group"
 [4] "returned_from_housing"
 [5] "returned_to_shelter"
 [6] "newly_identified"
 [7] "moved_to_housing"
 [8] "became_inactive"
 [9] "actively_homeless"
[10] "ageunder16"
[11] "age16_24"
[12] "age25_44"
[13] "age45_64"
[14] "age65over"
[15] "gender_male"
[16] "gender_female"
[17] "gender_transgender_non_binary_or_two_spirit"
[18] "population_group_percentage"
```

```
cleaned_shelter_data<-
  cleaned_shelter_data |>
  rename(
    dates = date_mmm_yy,
    group_percentage = population_group_percentage,
    trans_nb_twospirit = gender_transgender_non_binary_or_two_spirit,
    male = gender_male,
    female = gender_female,
    over65 = age65over,
    under16 = ageunder16
  )
head(cleaned_shelter_data)
```

```
# A tibble: 6 x 18
     id dates  populat~1 retur~2 retur~3 newly~4 moved~5 becam~6 activ~7 under16
  <dbl> <chr>  <chr>       <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1     1 Jan-18 All Popu~      46     494    1106     508     862    7958    1233
2     2 Jan-18 Chronic        11      29     317     111     111    2532     223
3     3 Jan-18 Refugees        4      32     651     328      75    2408     914
4     4 Jan-18 Families        0      14     561     321      57    2277    1232
5     5 Jan-18 Youth          12      64     116      44     116     924       0
6     6 Jan-18 Single A~      34     416     429     143     689    4757       0
# ... with 8 more variables: age16_24 <dbl>, age25_44 <dbl>, age45_64 <dbl>,
#   over65 <dbl>, male <dbl>, female <dbl>, trans_nb_twospirit <dbl>,
#   group_percentage <chr>, and abbreviated variable names 1: population_group,
#   2: returned_from_housing, 3: returned_to_shelter, 4: newly_identified,
#   5: moved_to_housing, 6: became_inactive, 7: actively_homeless
```

```
#### Simplifying the Dates: ####
cleaned_shelter_data <-
  cleaned_shelter_data |>
  separate(
    col = dates,
    into = c("Month", "Year"),
    sep = "-"
  )
head(cleaned_shelter_data)
```

```
# A tibble: 6 x 19
     id Month Year  population~1 retur~2 retur~3 newly~4 moved~5 becam~6 activ~7
```

```
  <dbl> <chr> <chr> <chr>          <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
1     1 Jan   18    All Populat~      46    494   1106    508    862   7958
2     2 Jan   18    Chronic          11    29    317    111    111   2532
3     3 Jan   18    Refugees          4    32    651    328     75   2408
4     4 Jan   18    Families          0    14    561    321     57   2277
5     5 Jan   18    Youth            12    64    116     44    116    924
6     6 Jan   18    Single Adult     34    416    429    143    689   4757
# ... with 9 more variables: under16 <dbl>, age16_24 <dbl>, age25_44 <dbl>,
#   age45_64 <dbl>, over65 <dbl>, male <dbl>, female <dbl>,
#   trans_nb_twospirit <dbl>, group_percentage <chr>, and abbreviated variable
#   names 1: population_group, 2: returned_from_housing,
#   3: returned_to_shelter, 4: newly_identified, 5: moved_to_housing,
#   6: became_inactive, 7: actively_homeless
```

```r
#### 2018: Average Number of Actively Homeless in Each Group ####
cleaned_shelter_data|>
  select(population_group, Year, Month, actively_homeless)|>
  filter(!population_group == "All Population")|>
  filter(Year == "18")|>
  group_by(population_group)|>
  summarise(mean = mean(actively_homeless))
```

```
# A tibble: 6 x 2
  population_group  mean
  <chr>            <dbl>
1 Chronic          2721.
2 Families         3050.
3 Non-refugees     5933.
4 Refugees         3238.
5 Single Adult     5205.
6 Youth             916.
```

```r
#### 2020: Average Number of Actively Homeless in Each Group ####
cleaned_shelter_data|>
  select(population_group, Year, Month, actively_homeless)|>
  filter(!population_group == "All Population")|>
  filter(Year == "20")|>
  group_by(population_group)|>
  summarise(mean = mean(actively_homeless))
```

```
# A tibble: 6 x 2
  population_group  mean
  <chr>            <dbl>
1 Chronic          3581.
2 Families         2034.
3 Non-refugees     6320.
4 Refugees         1921.
5 Single Adult     5423.
6 Youth             784.
```

```
#### 2022: Average Number of Actively Homeless in Each Group ####
cleaned_shelter_data|>
  select(population_group, Year, Month, actively_homeless)|>
  filter(!population_group == "All Population")|>
  filter(Year == "22")|>
  group_by(population_group)|>
  summarise(mean = mean(actively_homeless))
```

```
# A tibble: 7 x 2
  population_group  mean
  <chr>            <dbl>
1 Chronic          4803.
2 Families         2197.
3 Indigenous       1044.
4 Non-refugees     7185.
5 Refugees         2583.
6 Single Adult     6670.
7 Youth             902
```

Introduction

You can and should cross-reference sections and sub-sections. For instance, Section @ref(data). Quarto automatically makes the sections lower case and adds a dash to spaces to generate labels, for instance, Section @ref(first-discussion-point).
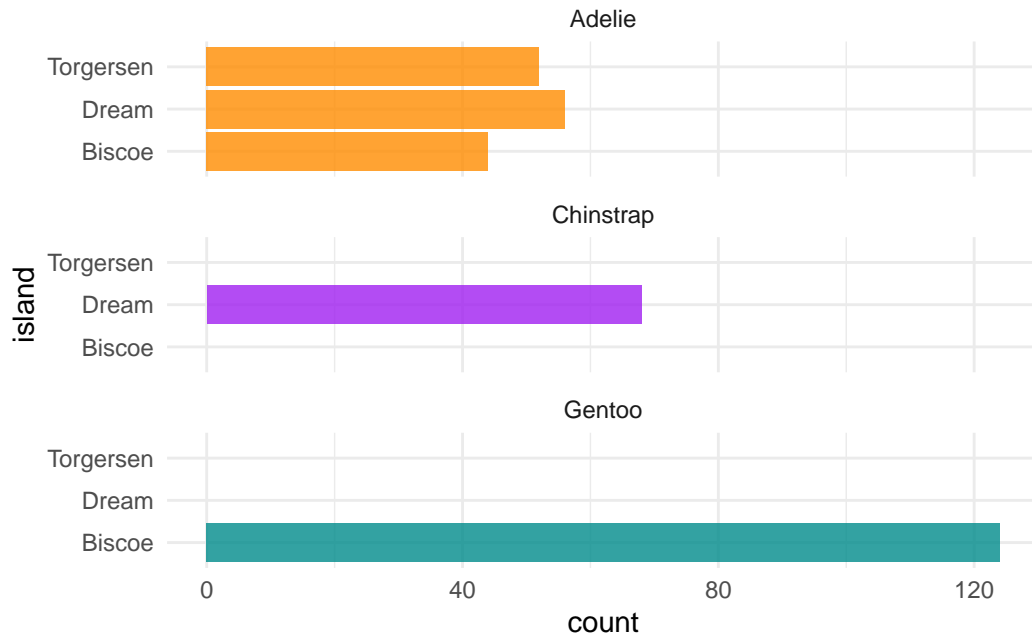
Data

Our data is of penguins (Figure 1).

Figure 1: Bills of penguins

Talk more about it.

Also bills and their average (Figure 2). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work.)

```
Warning: The `guide` argument in `scale_*()` cannot be `FALSE`. This was deprecated in
ggplot2 3.3.4.
i Please use "none" instead.
```

Talk way more about it.

Model

$$Pr(\theta|y) = \frac{Pr(y|\theta)Pr(\theta)}{Pr(y)} \tag{1}$$

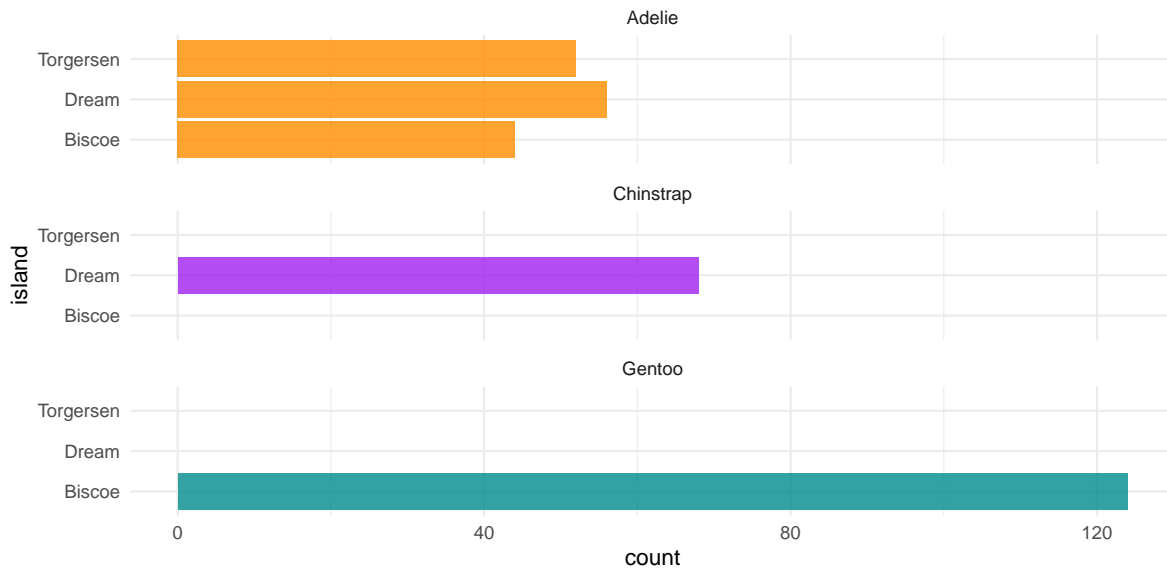Equation 1 seems useful, eh?

Figure 2: More bills of penguins

Here's a dumb example of how to use some references: In paper we run our analysis in R (R Core Team 2020). We also use the `tidyverse` which was written by Wickham et al. (2019) If we were interested in baseball data then Friendly et al. (2020) could be useful.

We can use maths by including latex between dollar signs, for instance θ.

```
?tibble
```

```
Help on topic 'tibble' was found in the following packages:

  Package               Library
  tidyr                 /Library/Frameworks/R.framework/Versions/4.2/Resources/library
  tibble                /Library/Frameworks/R.framework/Versions/4.2/Resources/library
  dplyr                 /Library/Frameworks/R.framework/Versions/4.2/Resources/library


Using the first match ...
```

Results

Discussion

First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Second discussion point

Third discussion point

Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

Additional details

References

Friendly, Michael, Chris Dalzell, Martin Monkman, and Dennis Murphy. 2020. *Lahman: Sean "Lahman" Baseball Database.* https://CRAN.R-project.org/package=Lahman.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.