

Hollywood Closet Project: *Exploratory Data Analysis*

Samuel Campione

- [Overview](#)
- [Research Questions](#)
- [1 Load the data set](#)
- [2 Data Wrangling and Cleaning](#)
- [3 Data Visualization and Summary](#)
- [3.1 Univariate Analysis](#)
- [3.2 Multivariate Analysis](#)

Overview

I worked on collecting this dataset as an RA in my undergrad. I decided to do a more in-depth analysis to see what insights may be gained about the data.

Because the sample was non-randomly selected and was not designed to comply with the theoretical assumptions of statistical tests, I decided to not include inferential procedures for lack of external validity. However, the descriptive statistics and visualizations provide fascinating insight into the sample. Further, these analyses provide an idea of how the population may look.

Future endeavors should focus on further collection of data to provide a more representative sample to gain firmer insight.

Research Questions

- What do the variables in this data set look like with descriptive statistics and data visualization?
 - How has the age of coming out changed over time?
 - Are people coming out at a younger age in more recent years?
 - How do all of these look across race, gender, and sexuality?
-

1 Load the data set

```
options(warn = -1) # Turn off warning messages

# Load required libraries
suppressMessages(library(tidyverse))
suppressMessages(library(wesanderson))
suppressMessages(library(forcats))
suppressMessages(library(knitr))
suppressMessages(library(kableExtra))
suppressMessages(library(ggrepel))
suppressMessages(library(patchwork))
```

1.1 Read in the file

```
data <- read.csv("HCP_data.csv", na.strings = c("", "NA"))
```

1.2 Preview the `data`

```
glimpse(data)
```

```
## Rows: 176
## Columns: 15
## $ ID <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
## $ First.Name <chr> "Jackie", "Ali", "James", "Roberto
## $ Last.Name <chr> "Abbott", "Adler", "Adomian", "Agu
## $ Race.Ethnicity <chr> "White", "White", "White", "Latinx
## $ Birth.Year <int> 1992, 1967, 1980, 1973, 1978, 1894
## $ Coming.Out.Age <int> 30, 20, 30, 31, 29, 27, 29, NA, NA
## $ Coming.Out.Year <int> 2017, 1987, 2010, 2004, 2007, 1921
## $ Coming.Out.Medium <chr> "Social Media", "Print", "Online",
## $ Type.of.Outing.Initiation <chr> "Other initiated", "Self initiated
## $ Type.of.Outing.Who.discloses. <chr> "Other's words", "Own words", "Oth
## $ Public.Sexual.Identity <chr> "Unspecified", "Lesbian", "Gay", "
## $ Were.they.in..the.closet.. <chr> "Unclear", "Unclear", "Yes", "No",
## $ Public.Gender.Identity <chr> "Cis Woman", "Cis Woman", "Cis Man
## $ X2nd.Coming.Out..Nature <chr> NA, "Amplified", NA, NA, NA, "Ampl
## $ X2nd.Coming.Out..Sexual.ID <chr> NA, "Lesbian", NA, NA, NA, "Lesbia
```

2 Data Wrangling and Cleaning

2.1 Relabel columns

```
colnames(data) = c("id", "first_name", "last_name", "race", "birth_year", "
  "coming_out_year", "medium", "initiation", "disclosure",
  "closeted", "gender", "coming_out_nature_2", "sexuality_2
```

2.2 Merge `first_name` and `last_name`

```
data <- data %>%
  unite("name", first_name:last_name, sep=" ")
```

2.3 Convert variables to correct data types

```
# All character variables to factors
data <- data %>%
  mutate_if(sapply(data, is.character), as.factor)
```

2.4 Create variable `coming_out_2` to easily explore subset

```
data<- data %>%
  mutate(coming_out_2 = ifelse(is.na(coming_out_nature_2), 0, 1) ,.after =
```

3 Data Visualization and Summary

3.1 Univariate Analysis

3.1.1 Race

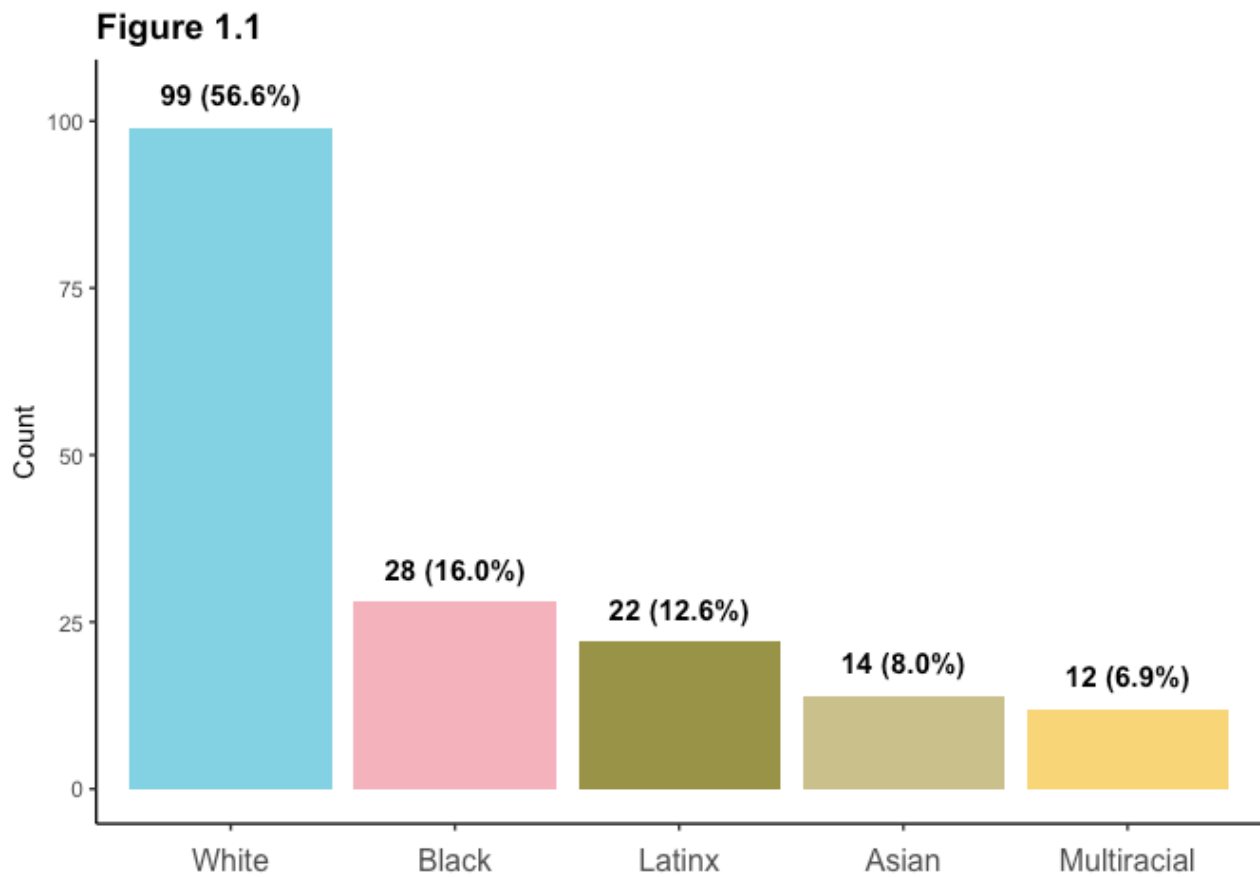


TABLE 1.1: Race

Race	Frequency	Relative Frequency
Asian	14	0.080
Black	28	0.159
Latinx	22	0.125
Multiracial	12	0.068
Unknown	1	0.006
White	99	0.562

Any missing values?

```
## [1] "Missing Values (NAs): 0"
```

3.1.2 Gender

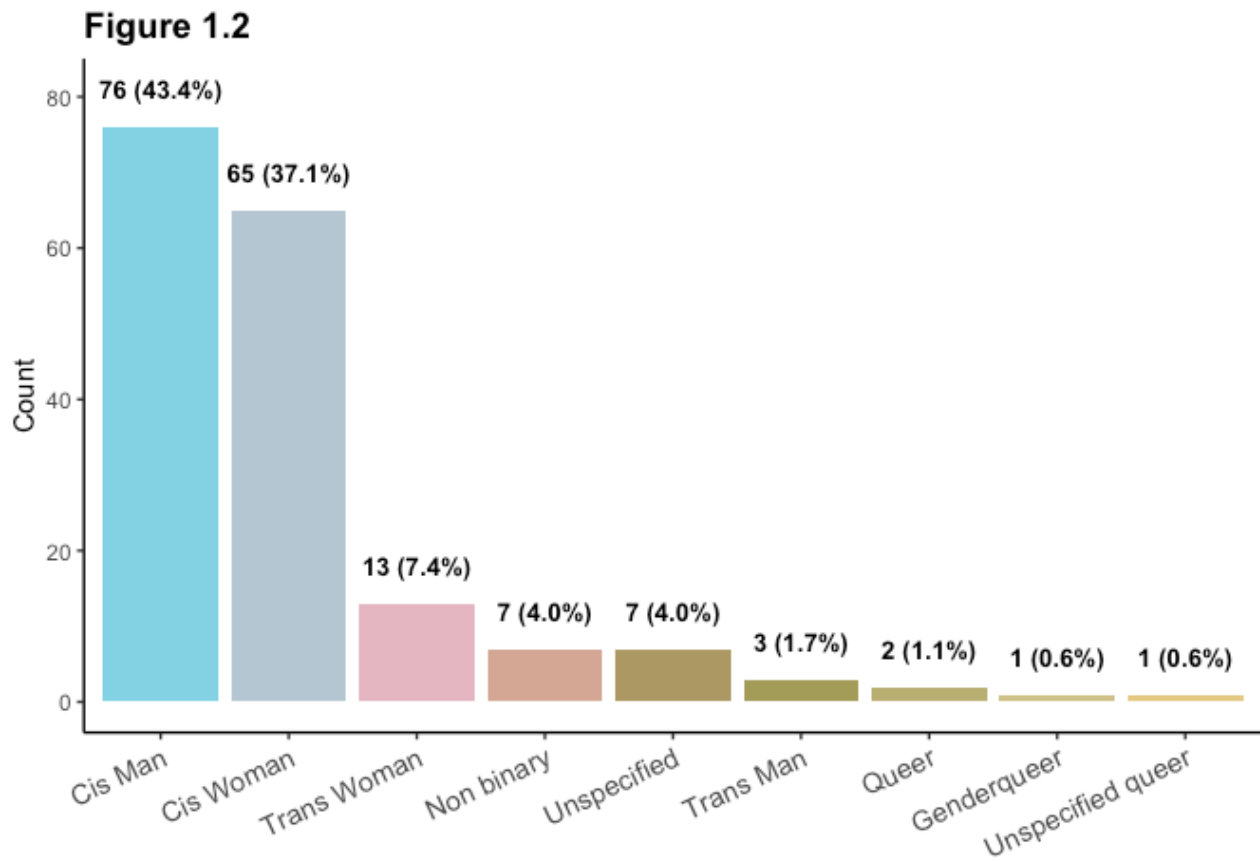


TABLE 1.2: Gender

Gender	Frequency	Relative Frequency
Cis Man	76	0.434
Cis Woman	65	0.371
Genderqueer	1	0.006
Non binary	7	0.040
Queer	2	0.011
Trans Man	3	0.017
Trans Woman	13	0.074
Unspecified	7	0.040
Unspecified queer	1	0.006

Any missing values?

```
## [1] "Missing Values (NAs): 1"
```

////////////////////////////////////

3.1.3 Sexuality

Figure 1.3

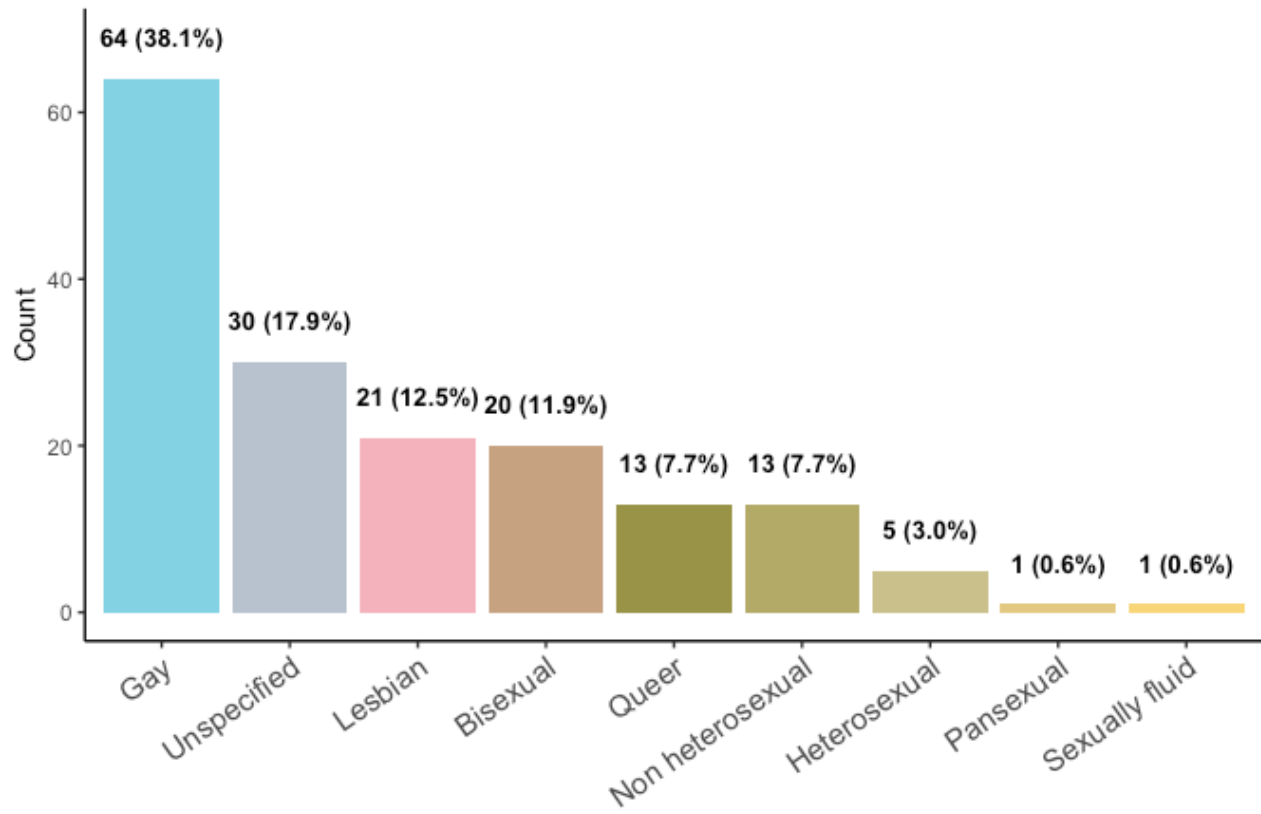


TABLE 1.3: Sexuality

Sexuality	Frequency	Relative Frequency
Bisexual	20	0.119
Gay	64	0.381
Heterosexual	5	0.030
Lesbian	21	0.125
Pansexual	1	0.006
Queer	13	0.077
Sexually fluid	1	0.006
Unspecified	30	0.179
Non heterosexual	13	0.077

Any missing values?

```
## [1] "Missing Values (NAs): 8"
```

////////////////////////////////////

3.1.2 In the closet?

Figure 1.4

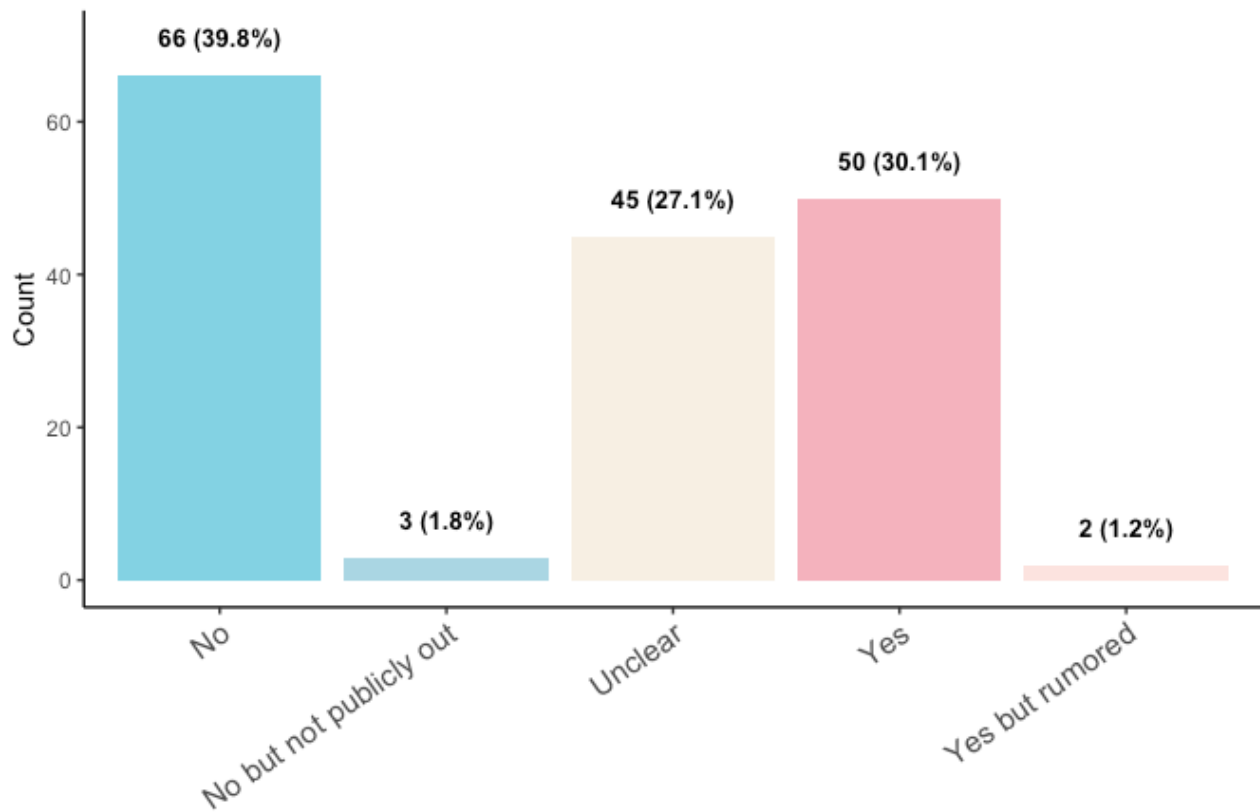


TABLE 1.4: Closeted before coming out

Closeted?	Frequency	Relative Frequency
No	66	0.398
No but not publicly out	3	0.018
Unclear	45	0.271
Yes	50	0.301
Yes but rumored	2	0.012

Any missing values?

```
## [1] "Missing Values (NAs): 10"
```

3.1.5 Coming Out Medium

Figure 1.5

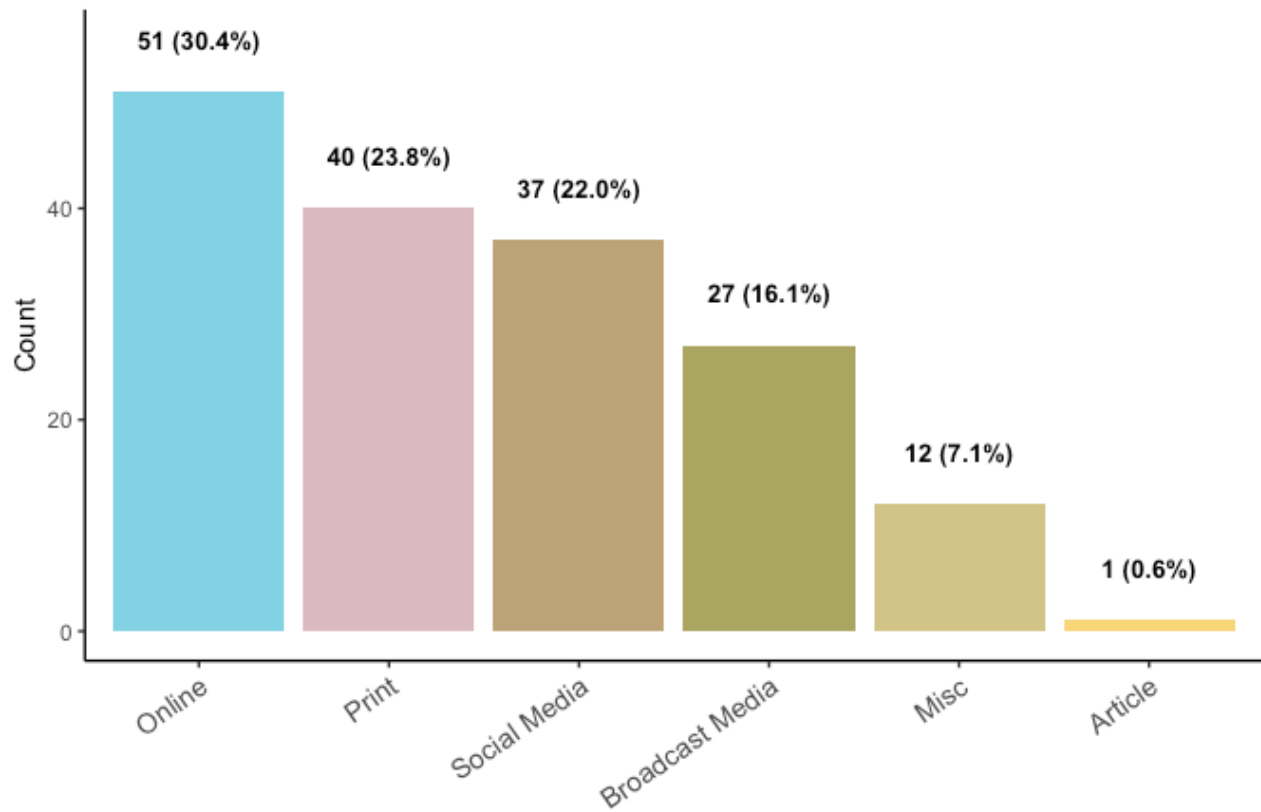


TABLE 1.5: Medium of coming out

Medium	Frequency	Relative Frequency
Article	1	0.006
Broadcast Media	27	0.161
Misc	12	0.071
Online	51	0.304
Print	40	0.238
Social Media	37	0.220

Any missing values?

```
## [1] "Missing Values (NAs): 8"
```

3.1.6 Who Initiated Coming Out?

A bar chart with 'Count' on the y-axis and two categories on the x-axis: 'Self initiated' and 'Other initiated'. The 'Self initiated' bar is light blue and reaches a count of 125, which is 75.3% of the total. The 'Other initiated' bar is light orange and reaches a count of 41, which is 24.7% of the total. The y-axis has major ticks at 0, 50, and 100.

Initiation Type	Count	Percentage
Self initiated	125	75.3%
Other initiated	41	24.7%

Initiation	Frequency	Relative Frequency
Other initiated	41	0.247
Self initiated	125	0.753

```
## [1] "Missing Values (NAs): 10"
```

3.1.7 Who Disclosed?

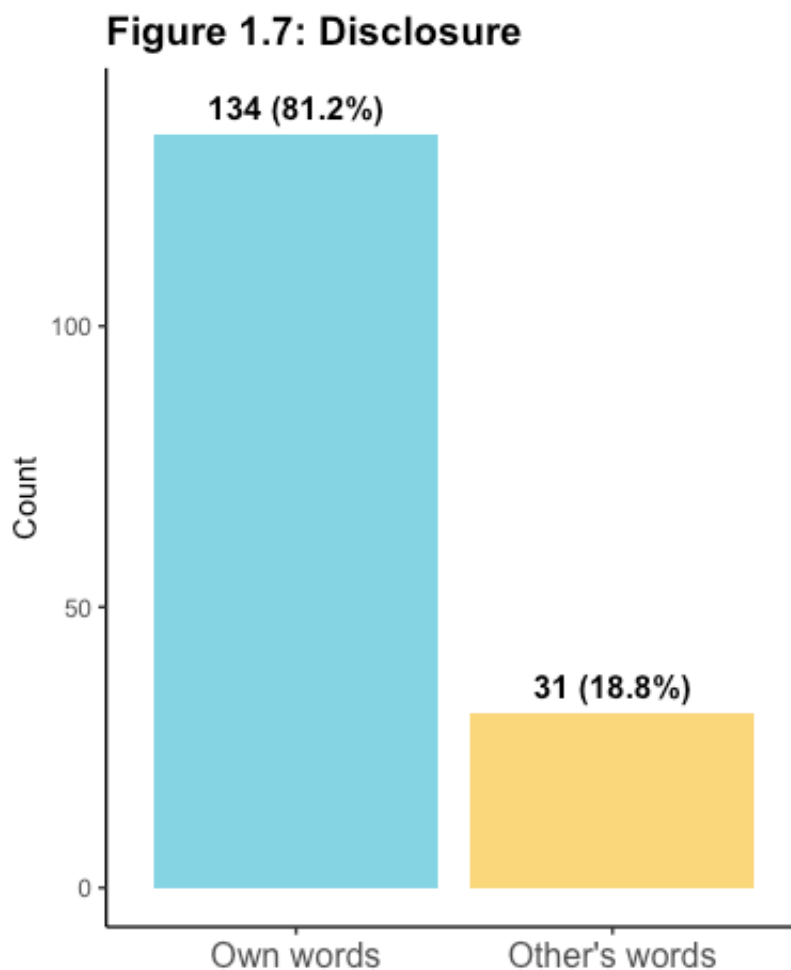


TABLE 1.7: Who disclosed

Disclosure	Frequency	Relative Frequency
Other's words	31	0.188
Own words	134	0.812

Any missing values?

```
## [1] "Missing Values (NAs): 11"
```

3.1.8 Had Second Coming Out?

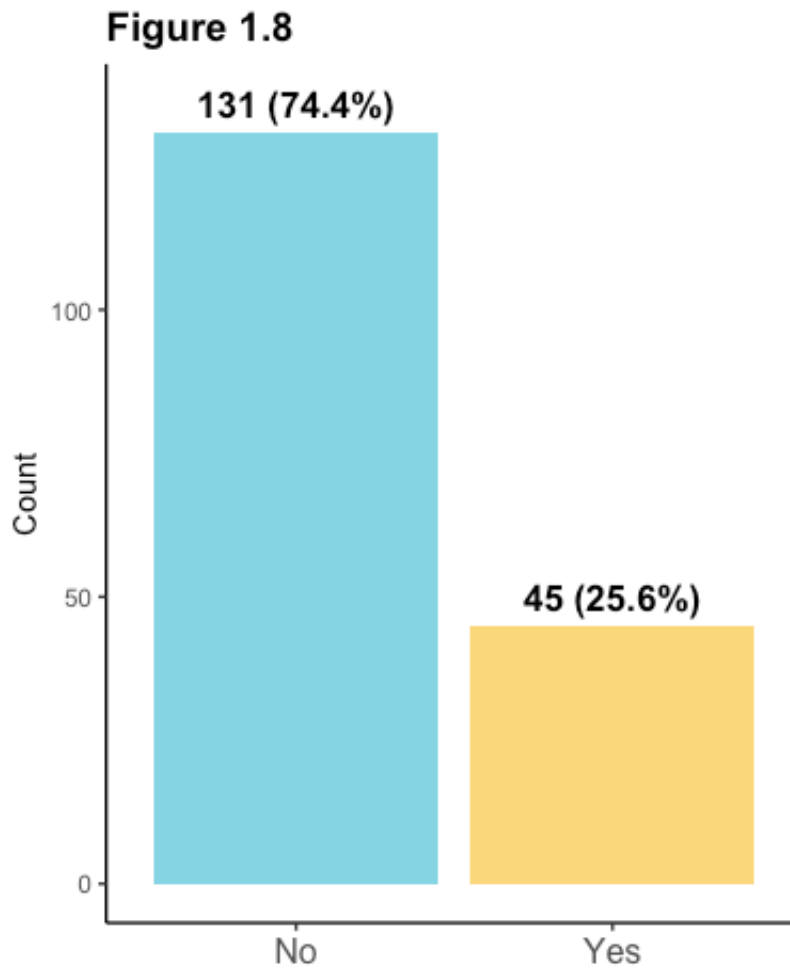


TABLE 1.8: Second coming out events

Had 2nd Coming Out	Frequency	Relative Frequency
No	131	0.74
Yes	45	0.26

Any missing values?

```
## [1] "Missing Values (NAs): 0"
```

3.1.9 Second Coming Out Nature

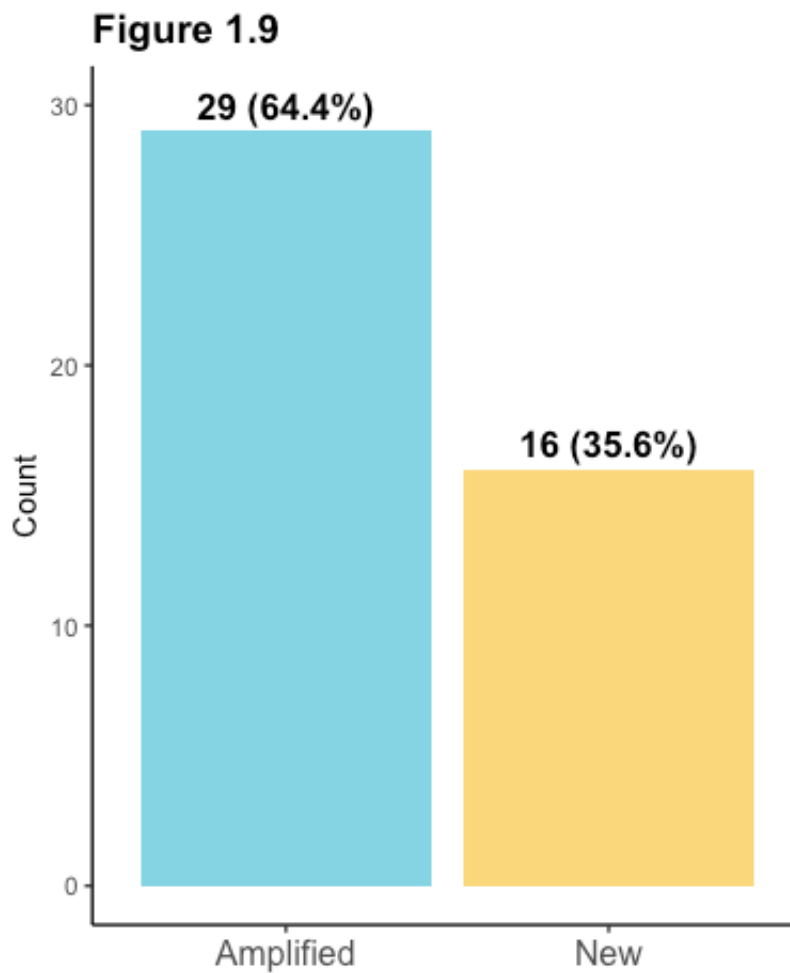


TABLE 1.9: Nature of second coming out

Nature	Frequency	Relative Frequency
Amplified	29	0.64
New	16	0.36

Any missing values?

```
## [1] "Missing Values (NAs): 0"
```

3.1.10 Second Coming Out Sexuality

Figure 1.10

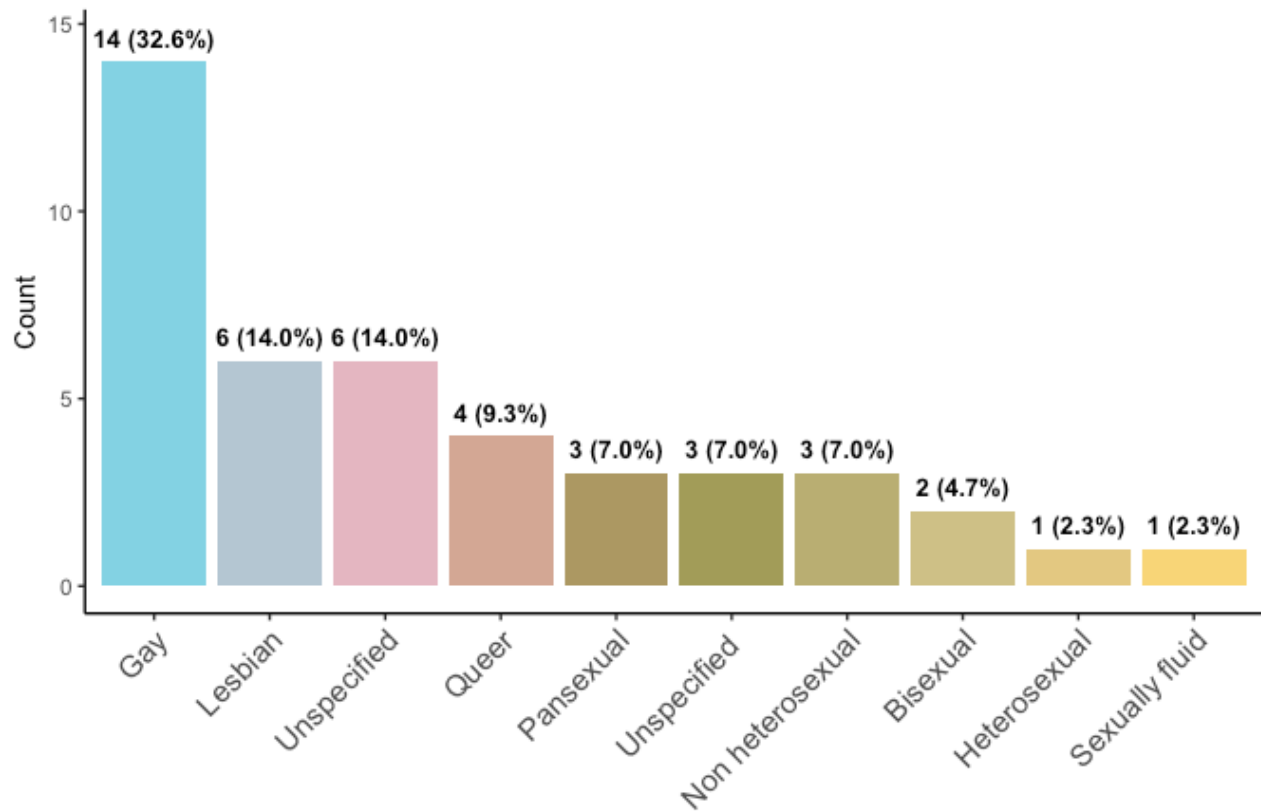


TABLE 1.10: Sexuality at second coming out

Sexuality	Frequency	Relative Frequency
Bisexual	2	0.05
Gay	14	0.33
Heterosexual	1	0.02
Lesbian	6	0.14
Pansexual	3	0.07
Queer	4	0.09
Sexually fluid	1	0.02
Unspecified	9	0.21
Non heterosexual	3	0.07

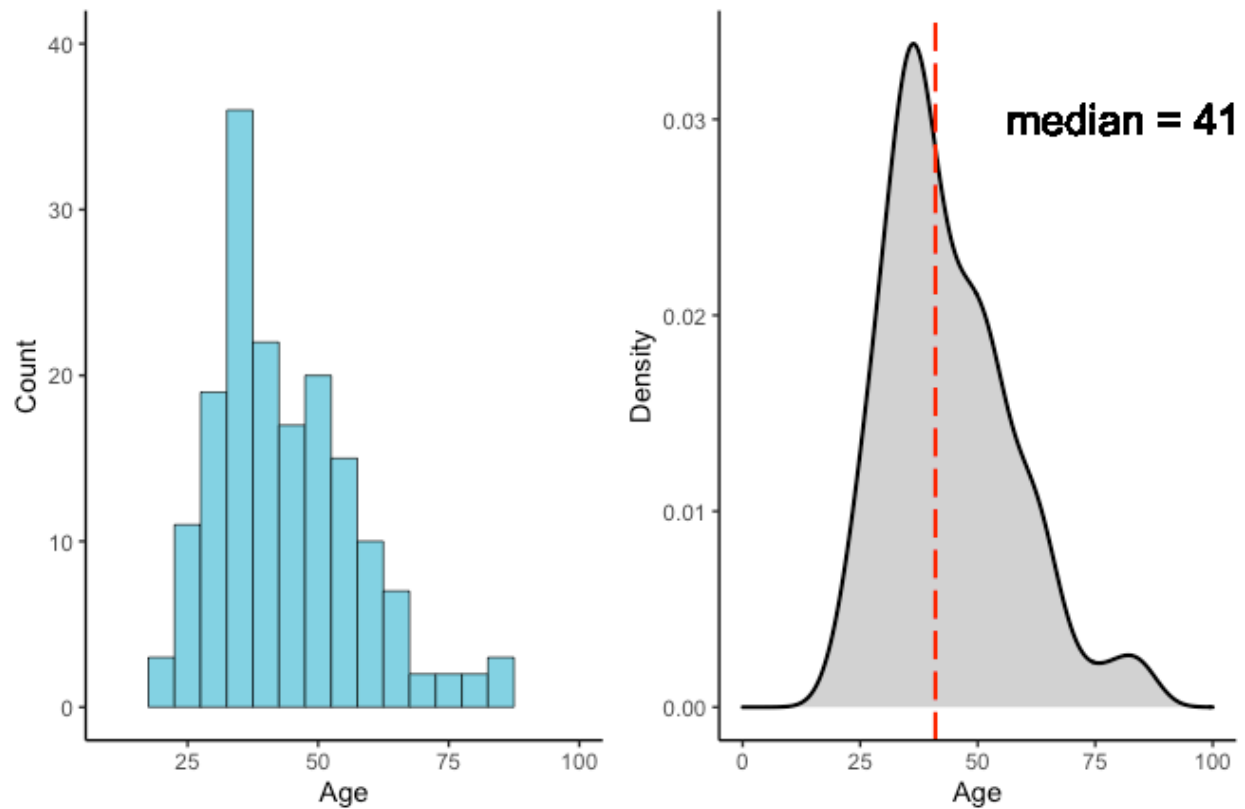
Any missing values?

```
## [1] "Missing Values (NAs): 2"
```

3.1.11 Current Age

Current ages in 2022 based on birth year. For simplicity, I applied a filter to include only those under 100 years old

Figure 1.11



Boxplot of Current Age

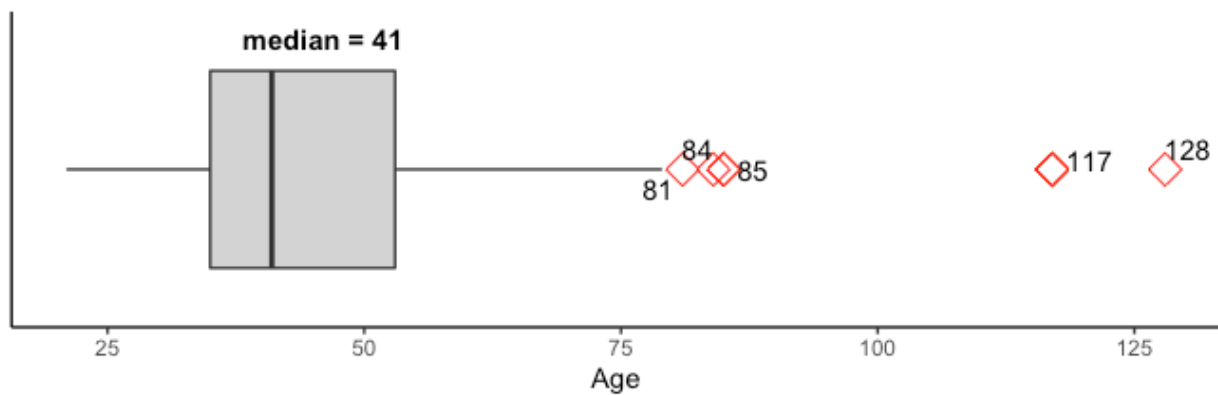
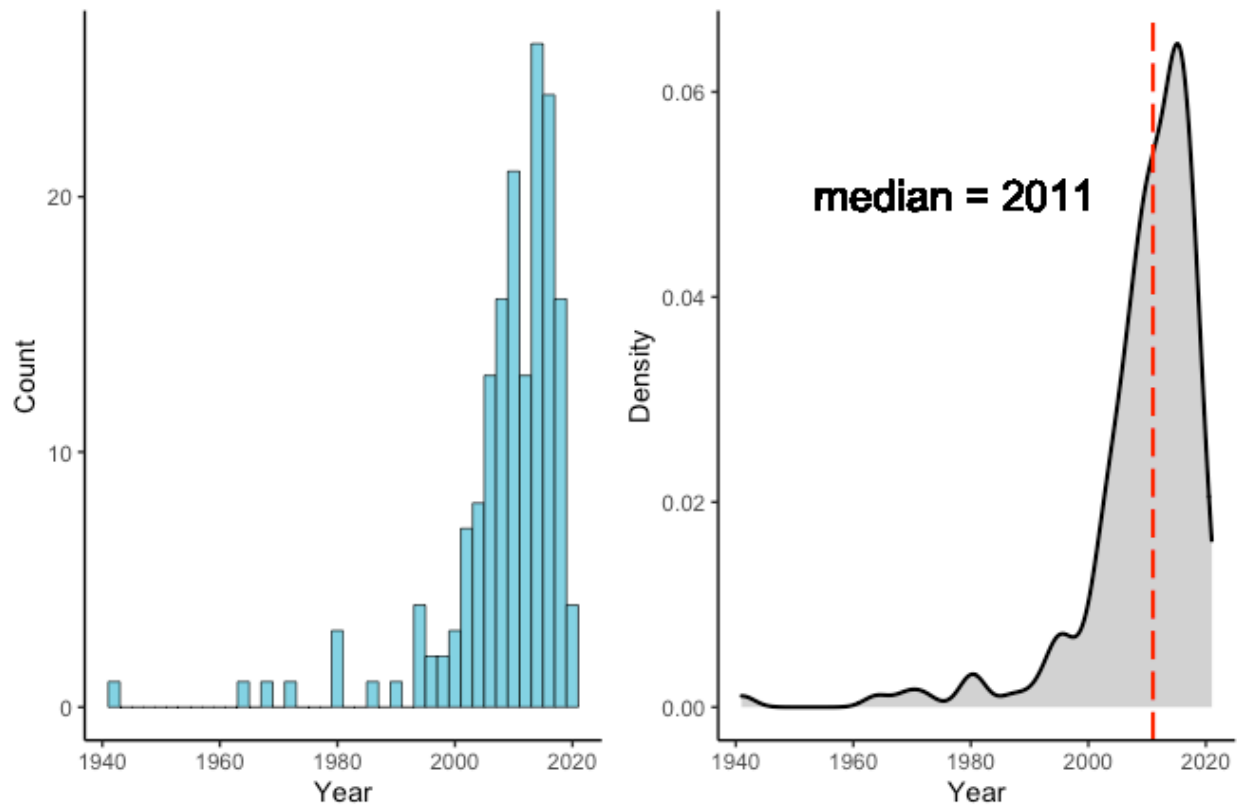


TABLE 1.11: Summary of Current Ages

Mean	SD	Median	Min	Max	NAs
45.6	17.6	41	21	128	3

3.1.12 Coming Out Year

Figure 1.12



Boxplot of Coming Out Year

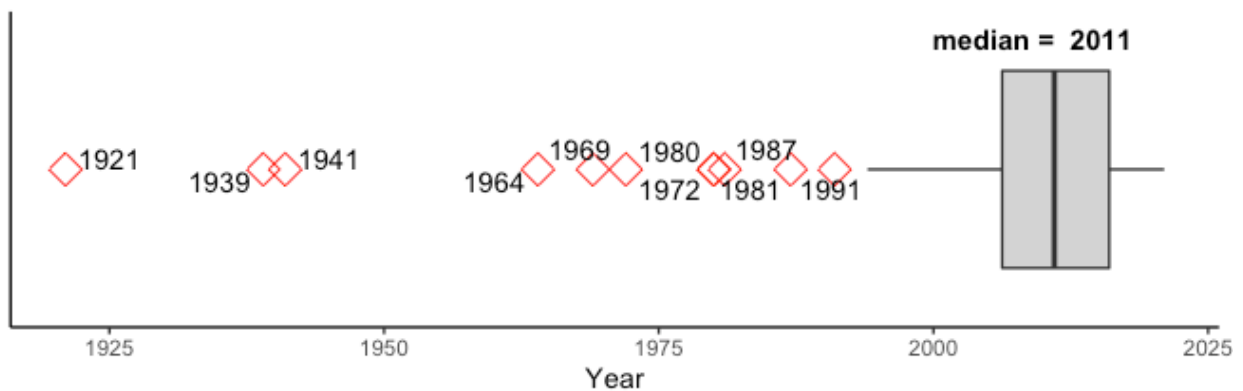
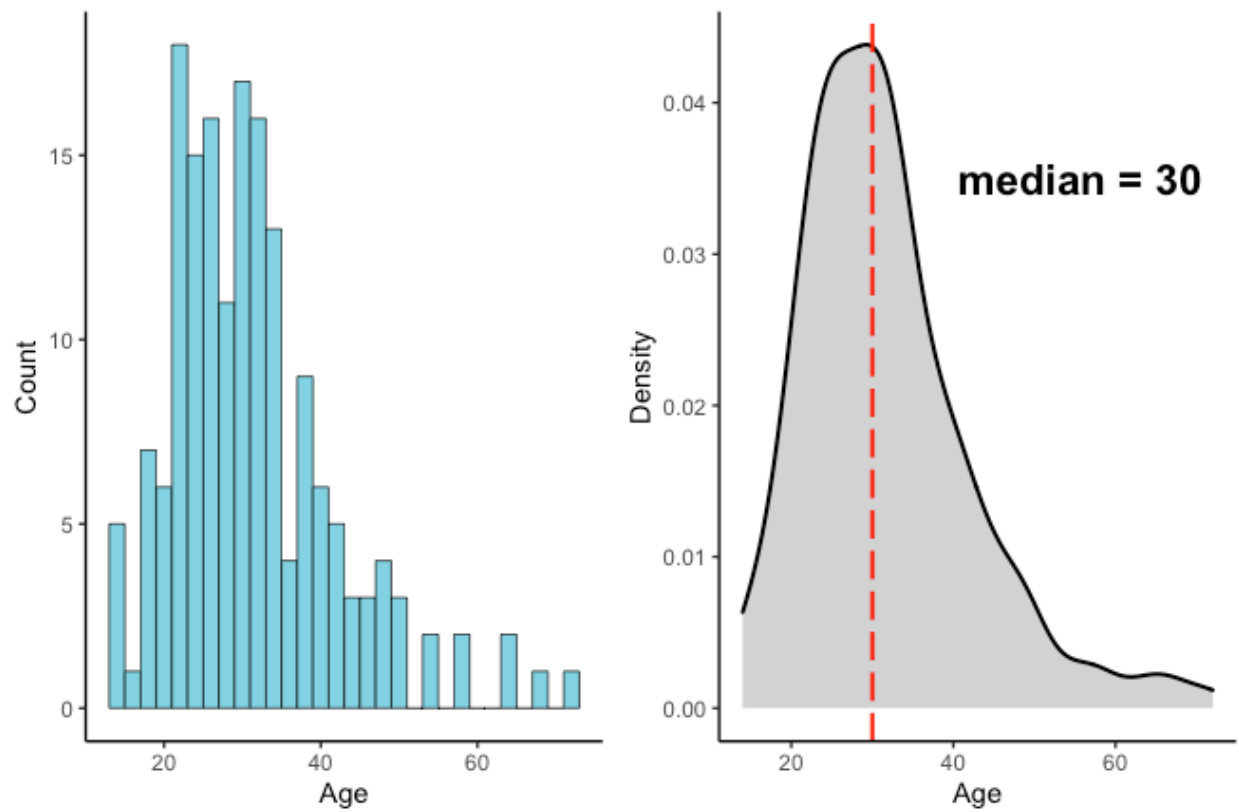


TABLE 1.12: Summary of Coming Out Years

Mean	SD	Median	Min	Max	NAs
2008	14	2011	1921	2021	0

3.1.13 Age at Coming Out

Figure 1.13



Boxplot of Coming Out Age

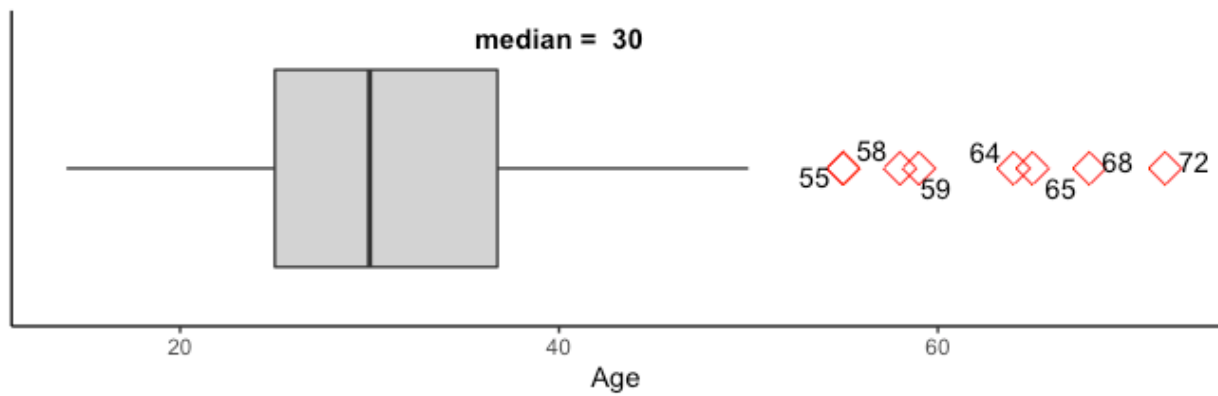


TABLE 1.13: Summary of Coming Out Age

Mean	SD	Median	Min	Max	NAs
31.8	10.6	30	14	72	0

3.2 Multivariate Analysis

3.2.1 Coming Out Age & Birth Year

Are younger generations coming out at a younger age?

Figure 2.1a

Coming Out Age and Generation

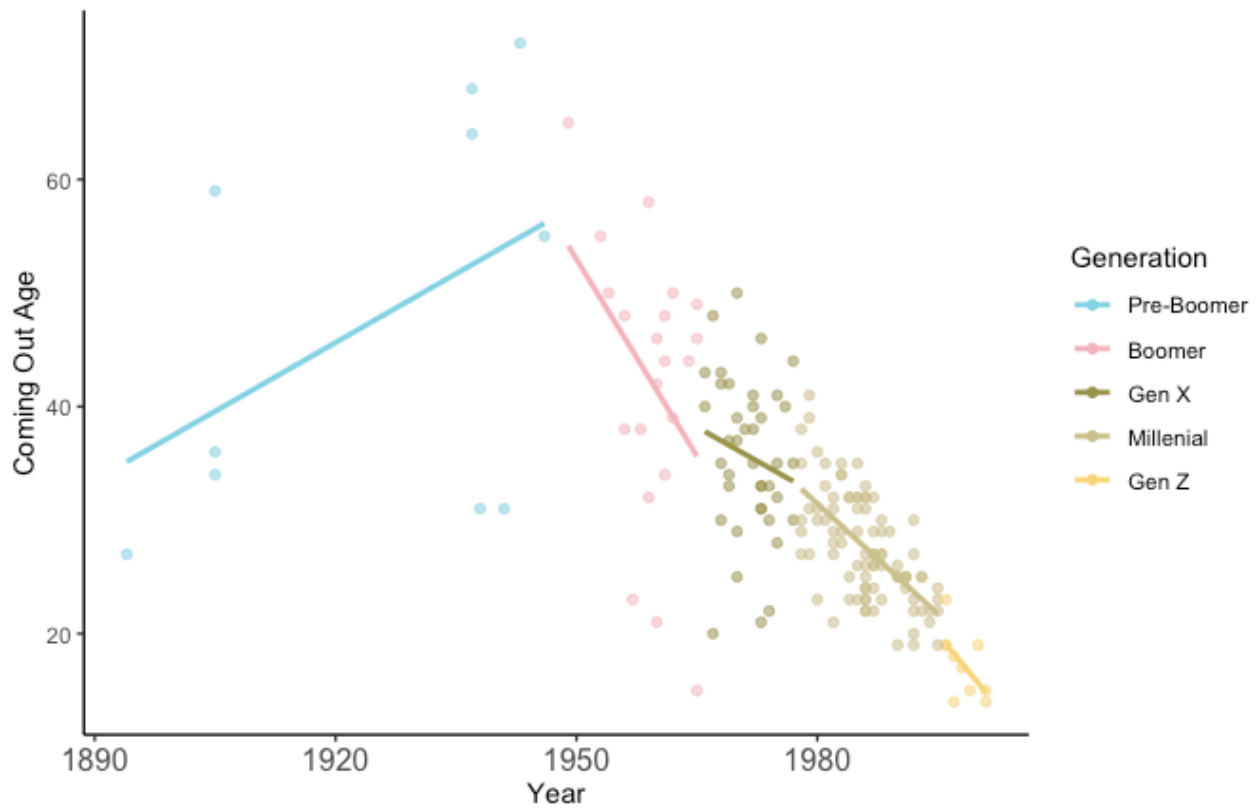


Figure 2.1b

Coming Out Age and Generation

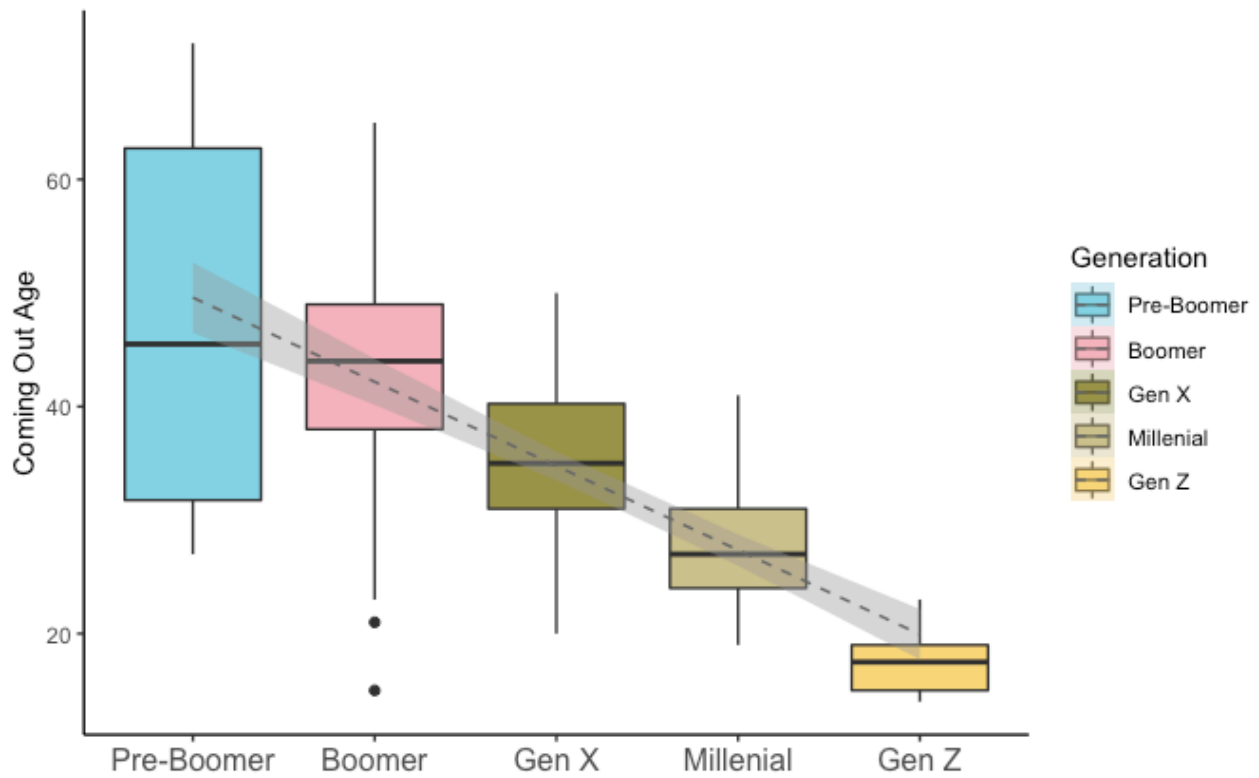


TABLE 2.1: Birth Decade and Coming Out Age

Decade	Count	Mean Coming Out Age
Pre-Boomer	10	47.70
Boomer	21	42.14
Gen X	40	35.58
Millennial	89	27.52
Gen Z	10	17.30

3.2.2. Coming Out Age & Race

What does coming out age look across race?

Figure 2.2

Coming Out Age by Race

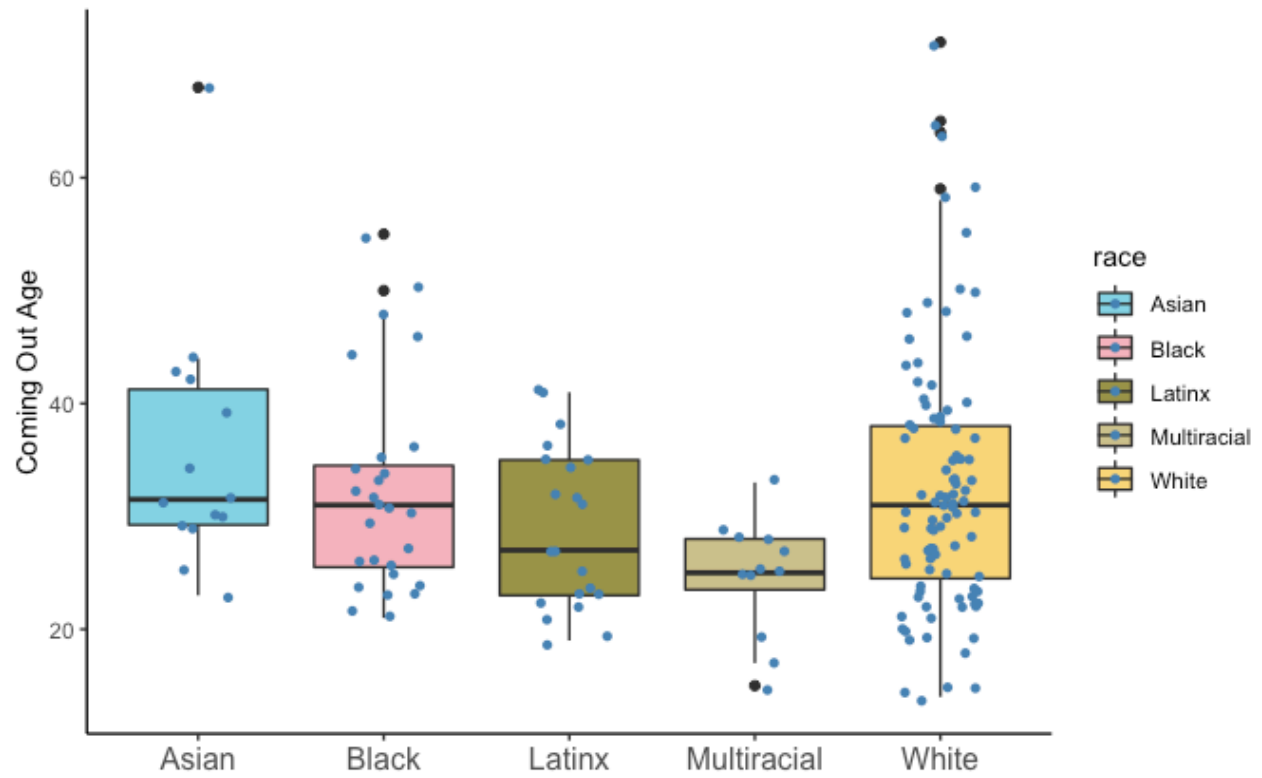


TABLE 2.2: Race and Coming Out Age

Race	Count	Mean	Median	Min	Max
Asian	14	35.64	31.5	23	68
Black	27	32.11	31.0	21	55
Latinx	21	28.90	27.0	19	41
Multiracial	12	24.67	25.0	15	33
Unknown	1	41.00	41.0	41	41
White	95	32.60	31.0	14	72

////////////////////////////////////

3.2.3 Coming Out Age & Gender

What does coming out age look across gender?

Figure 2.3

Coming Out Age by Gender

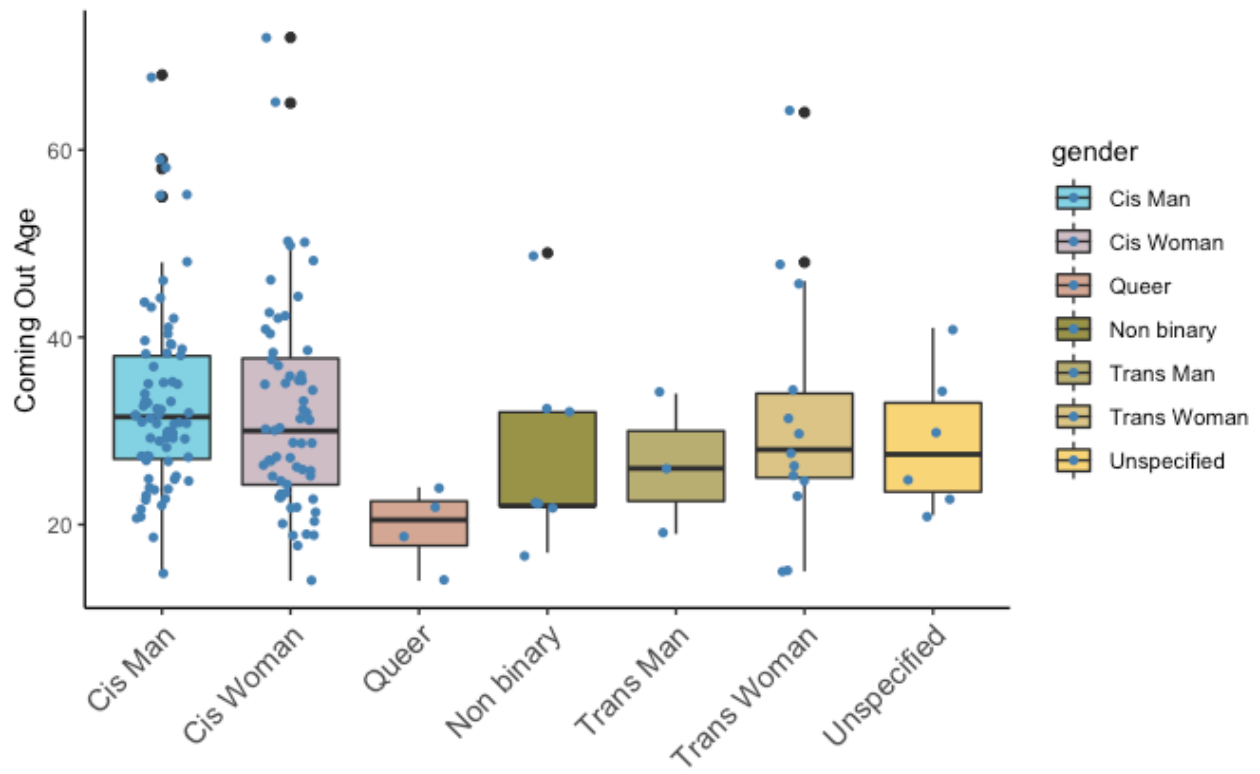


TABLE 2.3: Gender and Coming Out Age

Gender	Count	Mean	Median	Min	Max
Cis Man	74	33.16	31.5	15	68
Cis Woman	62	32.05	30.0	14	72
Queer	4	19.75	20.5	14	24
Non binary	7	28.00	22.0	17	49
Trans Man	3	26.33	26.0	19	34
Trans Woman	13	31.54	28.0	15	64
Unspecified	6	29.00	27.5	21	41



3.2.4 Coming Out Age & Sexuality

What does coming out age look across sexuality?

Figure 2.4

Coming Out Age by Sexuality

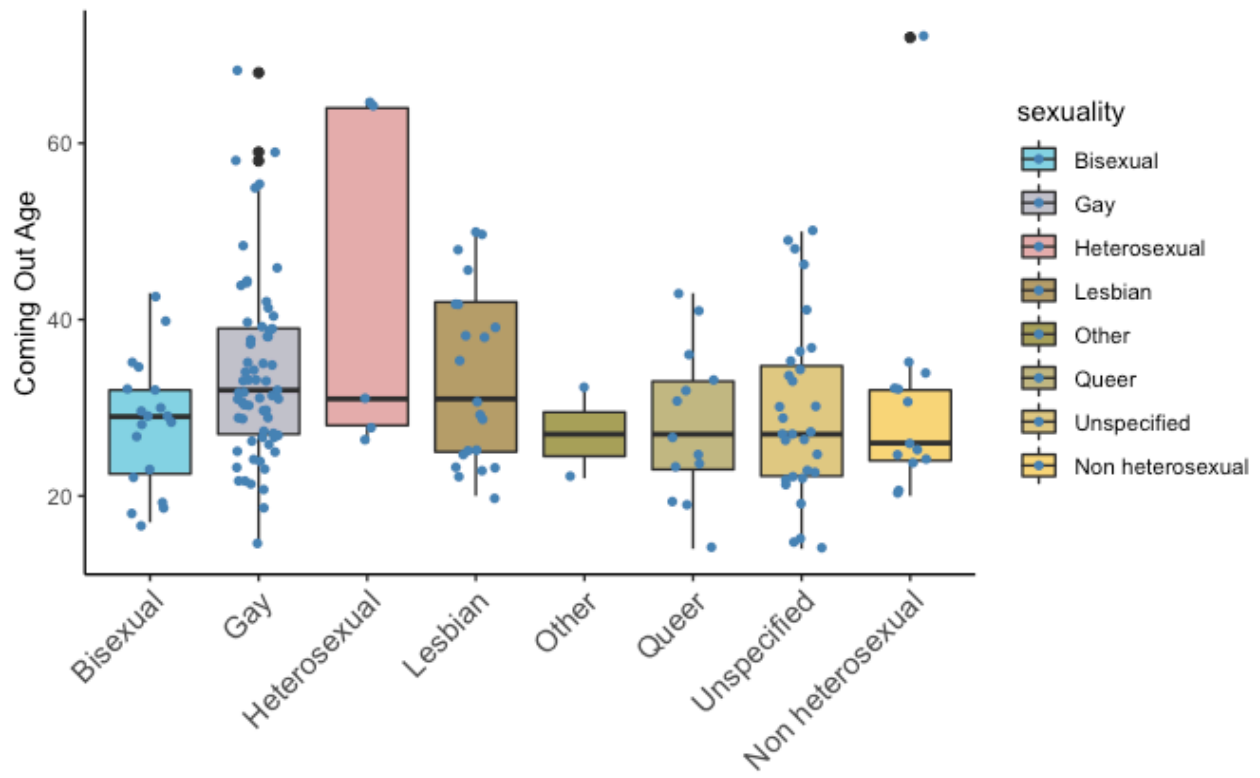


TABLE 2.4: Sexuality and Coming Out Age

Sexuality	Count	Mean	Median	Min	Max
Bisexual	19	28.21	29	17	43
Gay	64	33.84	32	15	68
Heterosexual	5	42.80	31	26	65
Lesbian	21	33.48	31	20	50
Pansexual	1	32.00	32	32	32
Queer	13	28.23	27	14	43
Sexually fluid	1	22.00	22	22	22
Unspecified	30	29.53	27	14	50
Non heterosexual	13	30.85	26	20	72