# Statistical analysis presentation

## Alcohol consumption in England (2011): an epidemiological perspective

Module 2 (Numerical Analysis) – PgDip in Artificial Intelligence

Guilherme Amorim

# Introduction

- Alcohol consumption landscape in England (2022):[1]
  - 81% of adults drink alcohol (55% at least weekly)
  - Increasing weekly average consumption (15.3 units/week in 2022, from 14.5 in 2021, and 10.9 in 2015)
  - Long-term risks (cardiovascular, gastrointestinal, cancer)[2]
- Assignment: explore demographic patterns related to alcohol consumption in England
- Dataset: Health Survey for England 2011
- Analysis plan: descriptive and inferential statistics
- Analysis tools: R

[1]NHS England (2022) *Health Survey for England*. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england.

[2]NHS (2022) 'Risks - Alcohol misuse'. Available from: https://www.nhs.uk/conditions/alcohol-misuse/risks/.

# Descriptive statistics

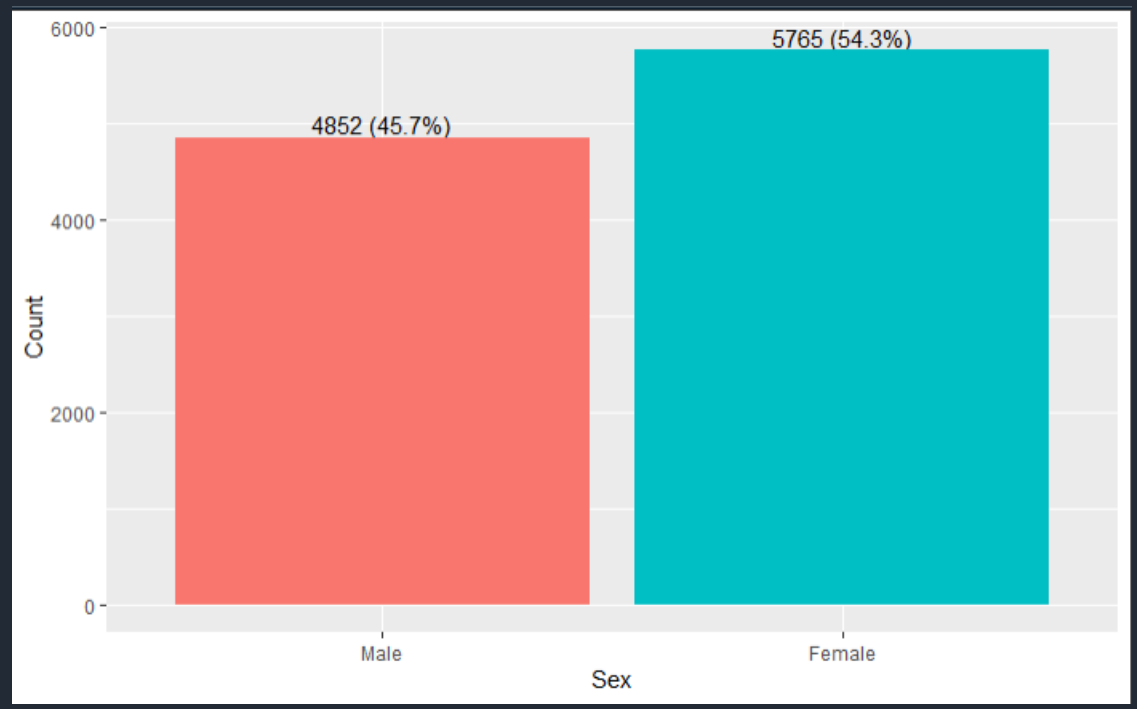
# Descriptive statistics

Q1: How many people are included in the sample?

10617

Q2: What is the percentage of women in the sample?

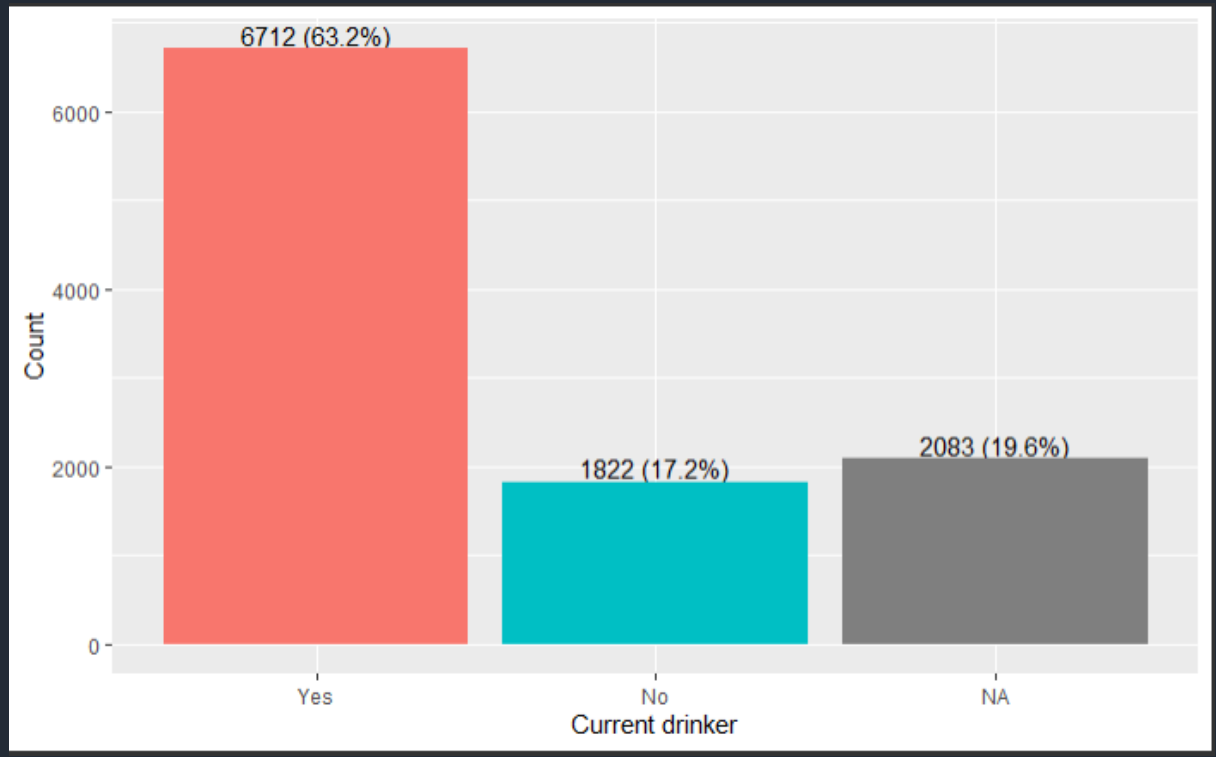
54.3% women (5765)



# Descriptive statistics

Q3: What is the percentage of people who drink alcohol?

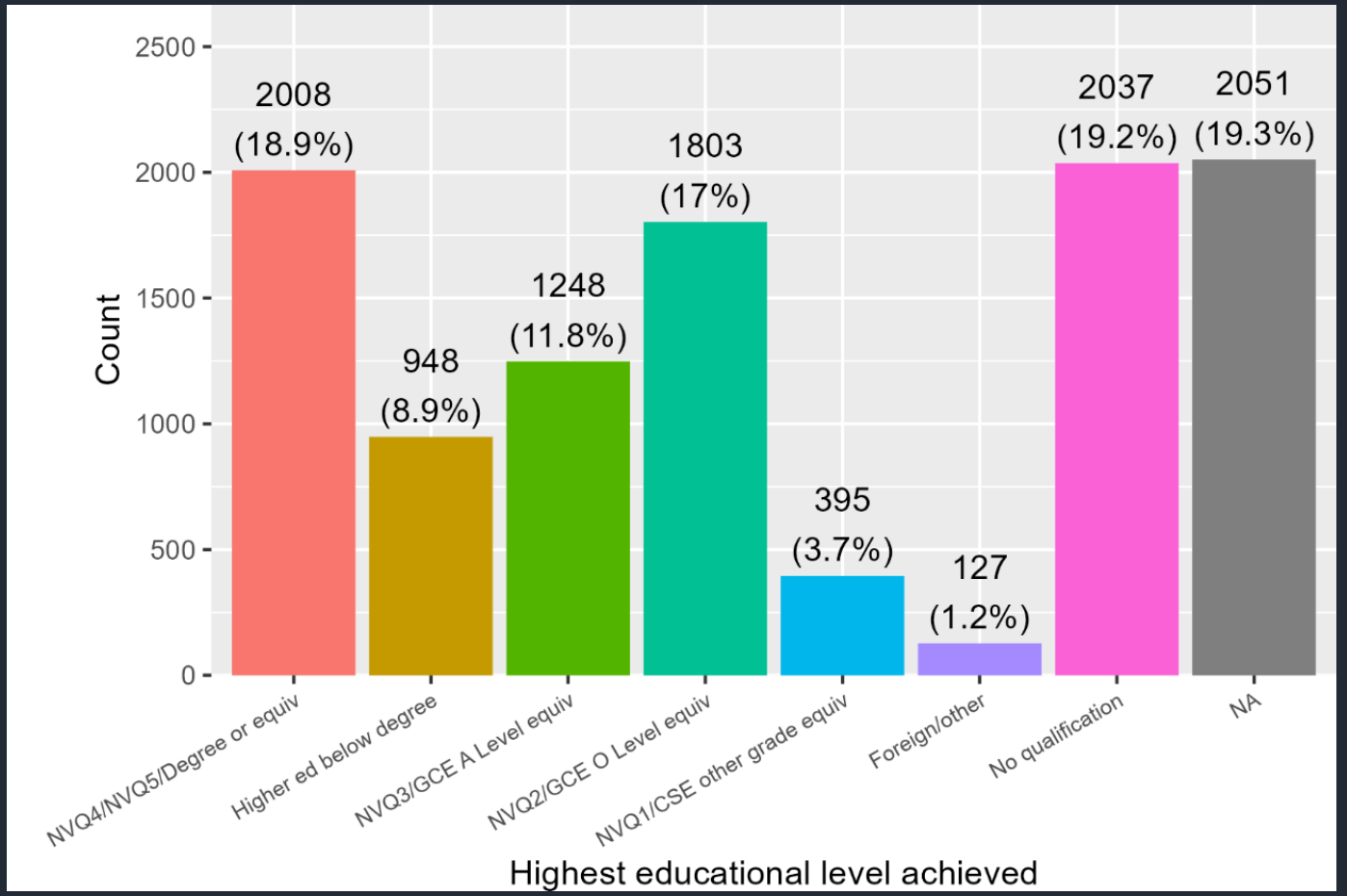
6712 drink alcohol nowadays (63.2%; 78.7% if excluding those with missing answers)



# Descriptive statistics

Q4: What is the highest educational level?

NVQ4/NVQ5/Degree or equivalent (18.9%, 2008 people)

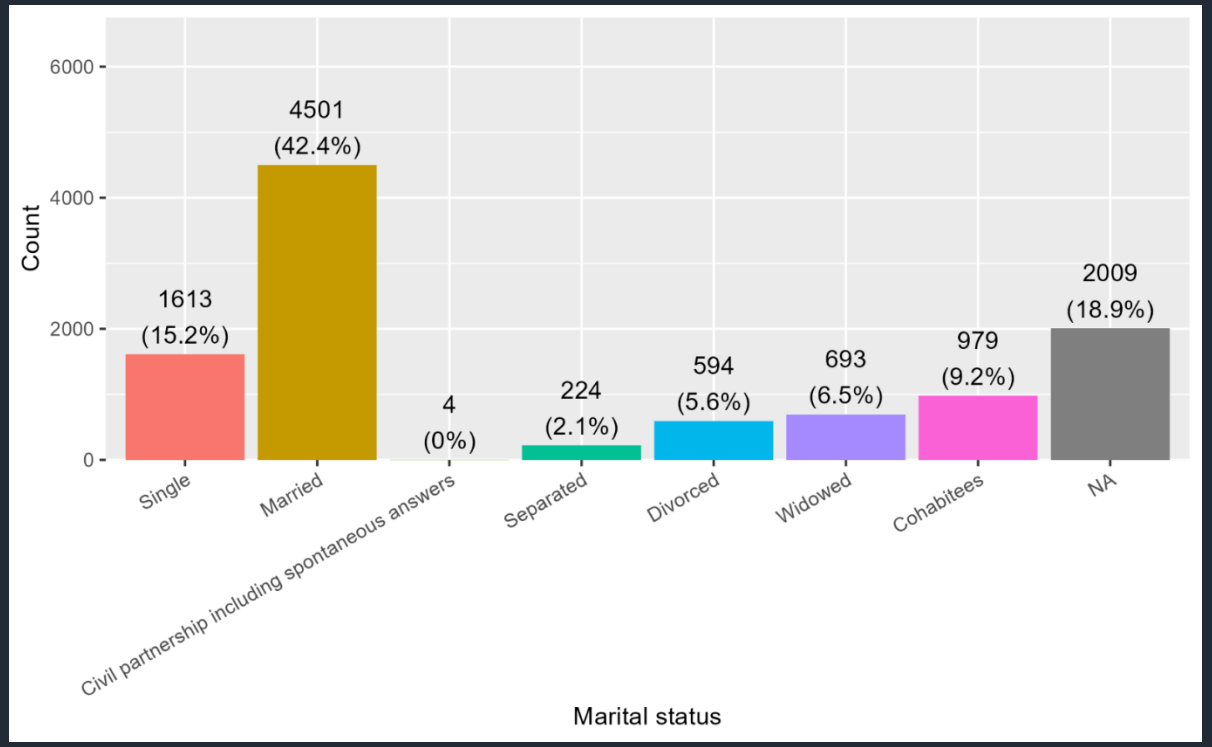


# Descriptive statistics

Q5: What is percentage of divorced and separated people?

Divorced: 594
(5.6%; 6.9% if excluding those with missing data)

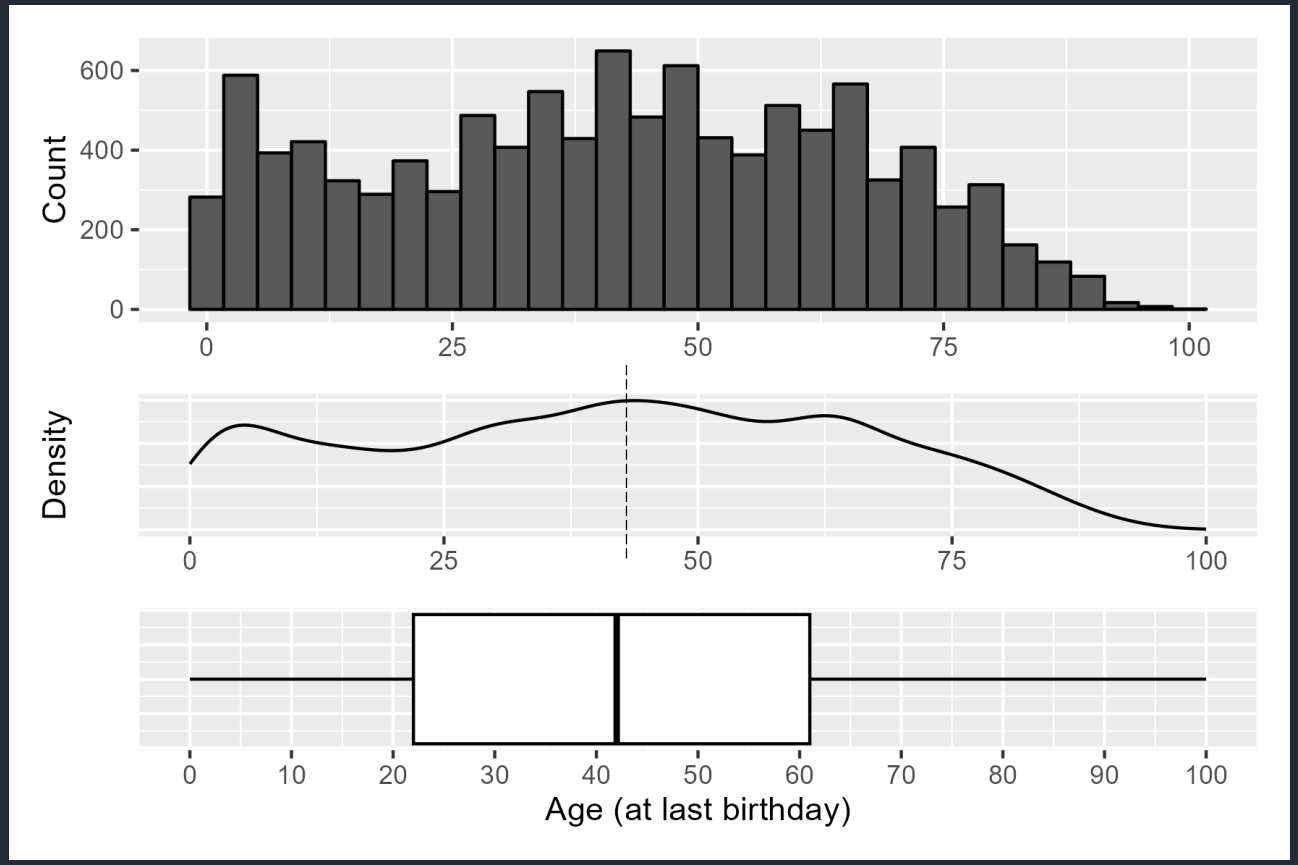
Separated: 224
(2.1%; 2.6% if excluding those with missing data)



# Descriptive statistics

Q6.1: Find the mean, median, mode, minimum, maximum, range and standard deviation of age at last birthday

| Statistic | Estimate |
|---|---|
| Mean | 41.6 |
| Median | 42 |
| Mode | 43.77* |
| Minimum | 0 |
| Maximum | 100 |
| Range | 0 |
| Standard deviation | 23.8 |



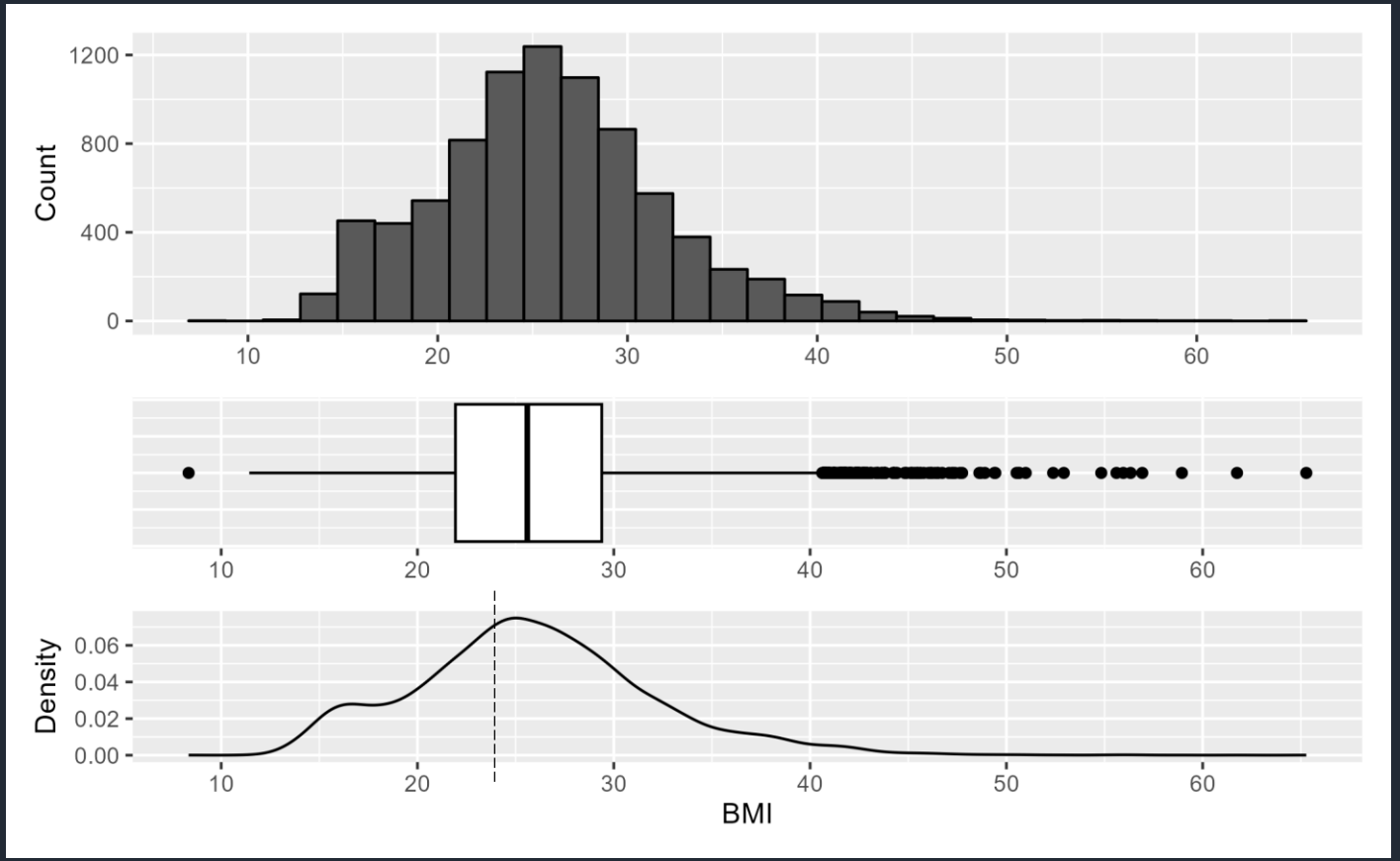

* Calculated from the maximum of the probability density function (dashed line on density plot)

# Descriptive statistics

Q6.2: Find the mean, median, mode, minimum, maximum, range and standard deviation of BMI

| Statistic | Estimate |
|---|---|
| Mean | 25.9 |
| Median | 25.59 |
| Mode | 25* |
| Minimum | 8.34 |
| Maximum | 65.28 |
| Range | 56.94 |
| Standard deviation | 6.14 |

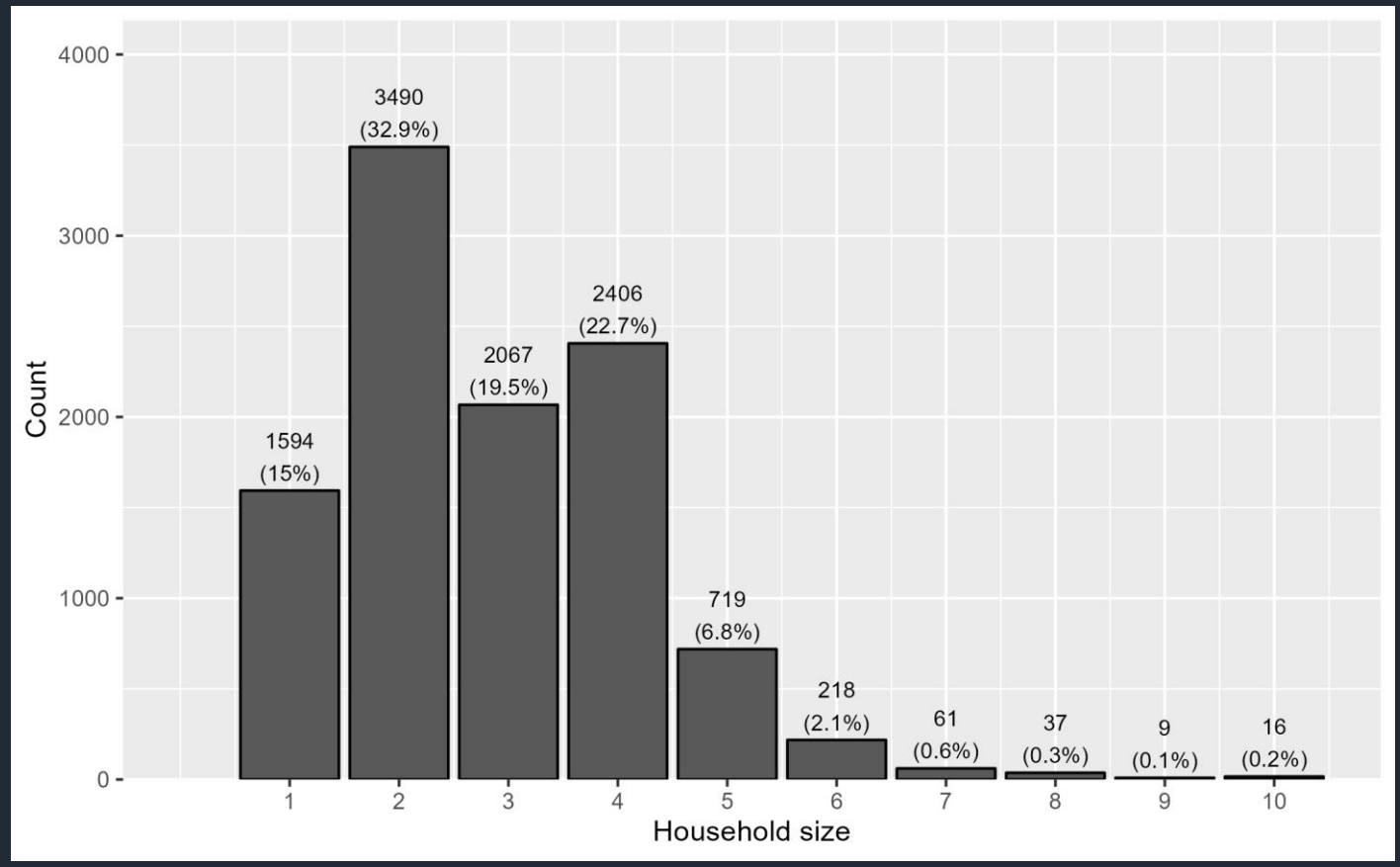
* Calculated from the maximum of the probability density function (dashed line on density plot)



# Descriptive statistics

Q6.3: Find the mean, median, mode, minimum, maximum, range and standard deviation of household size

| Statistic | Estimate |
|---|---|
| Mean | 2.85 |
| Median | 3 |
| Mode | 2 |
| Minimum | 1 |
| Maximum | 10 |
| Range | 9 |
| Standard deviation | 1.37 |



# Inferential statistics

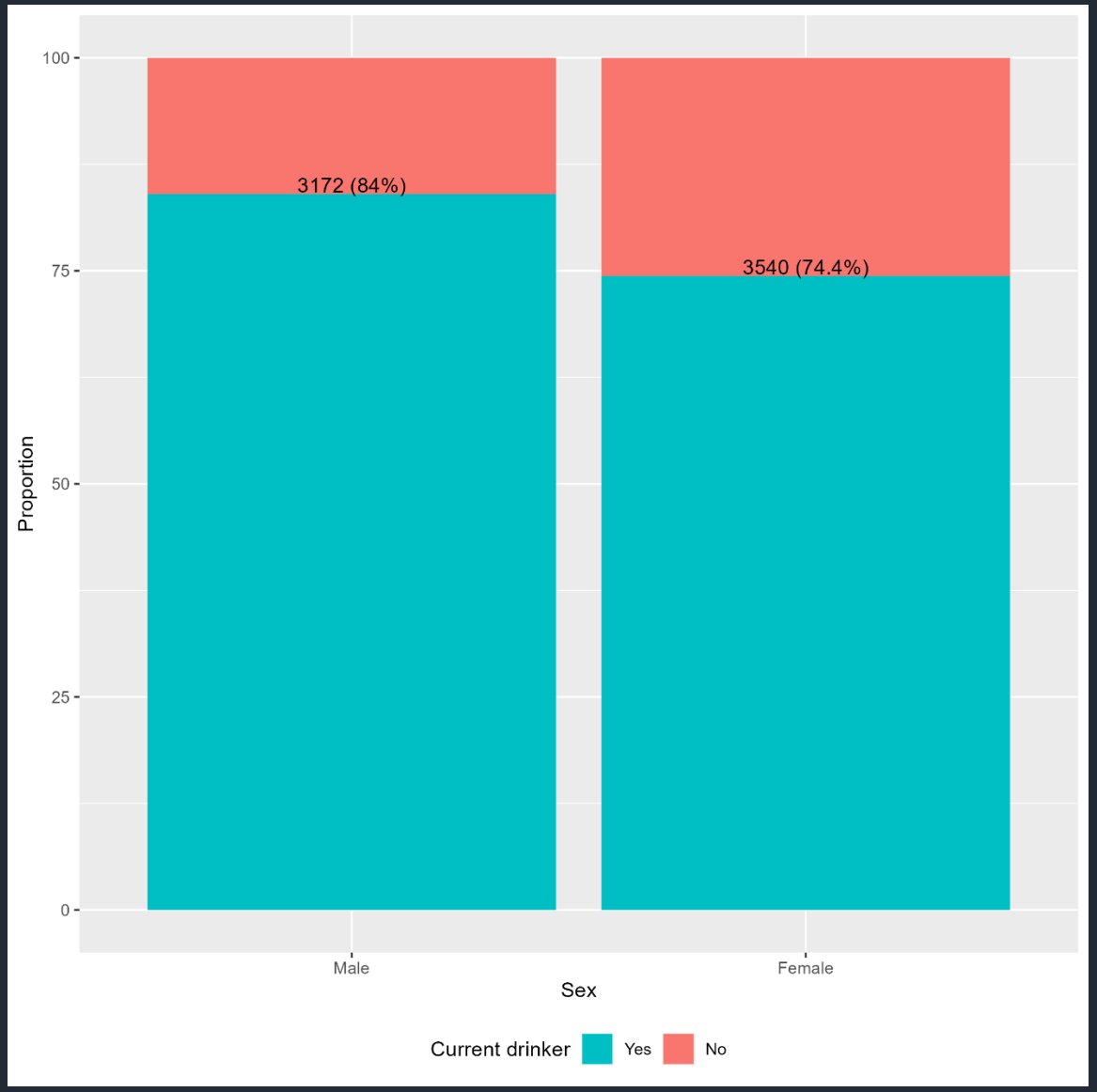
# Inferential statistics

Q7: Which sex drinks more alcohol?

| | | Current drinker | | Test statistic and p-value |
|---|---|---|---|---|
| | | Yes | No | |
| Sex | Male | 3172 (84%) | 1650 (16%) | $\chi 2$= 114.15<br>P-value<0.001 |
| | Female | 3540 (74.4%) | 2225 (25.5%) | |

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  data$drinks and data$sex
X-squared = 114.15, df = 1, p-value < 2.2e-16
```



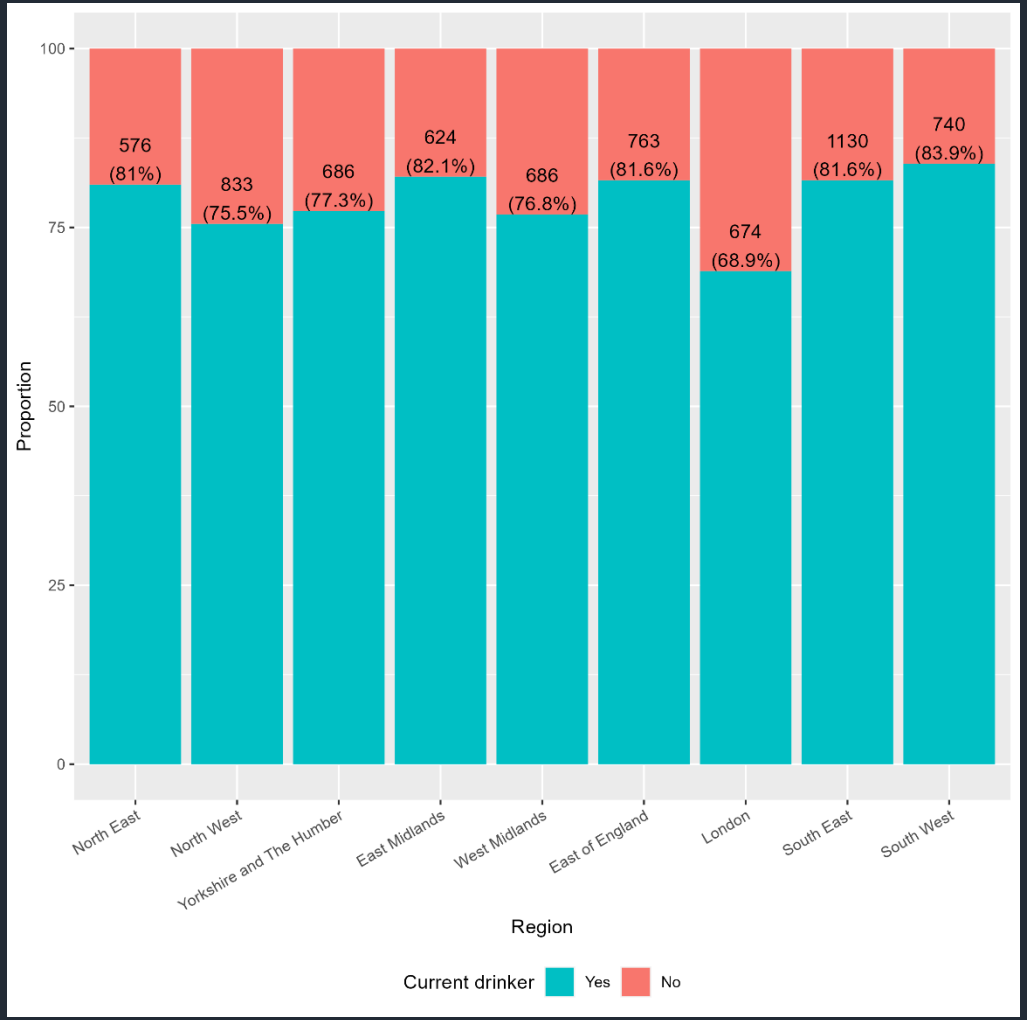
# Inferential statistics

## Q8: Which region drinks the most alcohol?

| | | Current drinker | | Test statistic and p-value |
|---|---|---|---|---|
| | | Yes | No | |
| Region | North East | 576 (81.01%) | 135 (18.99%) | $\chi^2$= 98.53<br>P-value<0.001 |
| | North West | 833 (75.52%) | 270 (24.48%) | |
| | Yorkshire and The Humber | 686 (77.34%) | 201 (22.66%) | |
| | East Midlands | 624 (82.11%) | 136 (17.89%) | |
| | West Midlands | 686 (76.82%) | 207 (23.18%) | |
| | East of England | 763 (81.60%) | 172 (18.40%) | |
| | London | 674 (68.92%) | 304 (31.08%) | |
| | South East | 1130 (81.59%) | 255 (18.41%) | |
| | South West | 740 (83.90%) | 142 (16.10%) | |

```
Pearson's Chi-squared test

data:  data$drinks and data$region
X-squared = 98.53, df = 8, p-value < 2.2e-16
```



# Inferential statistics

Q9: Is there a statistical difference on valid height between men and women?

| Sex | Min | Mean | SD | Median | Q1 | Q3 | Max |
|---|---|---|---|---|---|---|---|
| Men | 84.8 | 167.4 | 21 | 173.3 | 166.5 | 179 | 202.5 |
| Women | 82.4 | 157.2 | 15.4 | 160.4 | 154.9 | 165.6 | 186.4 |

Non-normally distributed variable (K-S test)

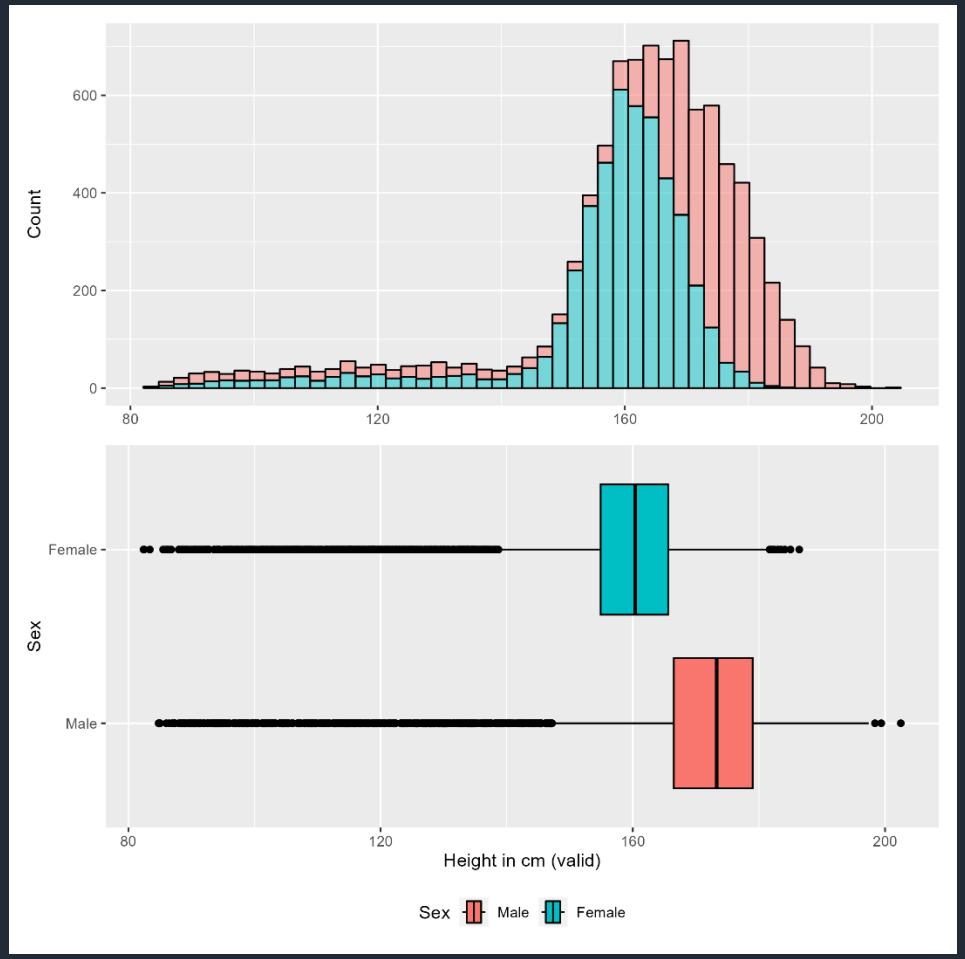
```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  data$height
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Non-parametric Wilcoxon independent samples test

```
        Wilcoxon rank sum test with continuity correction

data:  height by sex
W = 14713021, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```



# Inferential statistics

Q10: Is there a statistical difference on valid weight between men and women?

| Sex | Min | Mean | SD | Median | Q1 | Q3 | Max |
|---|---|---|---|---|---|---|---|
| Men | 4.6 | 74.3 | 27 | 78.8 | 65.2 | 90 | 184.3 |
| Women | 1 | 64.8 | 22 | 65.7 | 55.5 | 77.3 | 172 |

Non-normally distributed variable (K-S test)

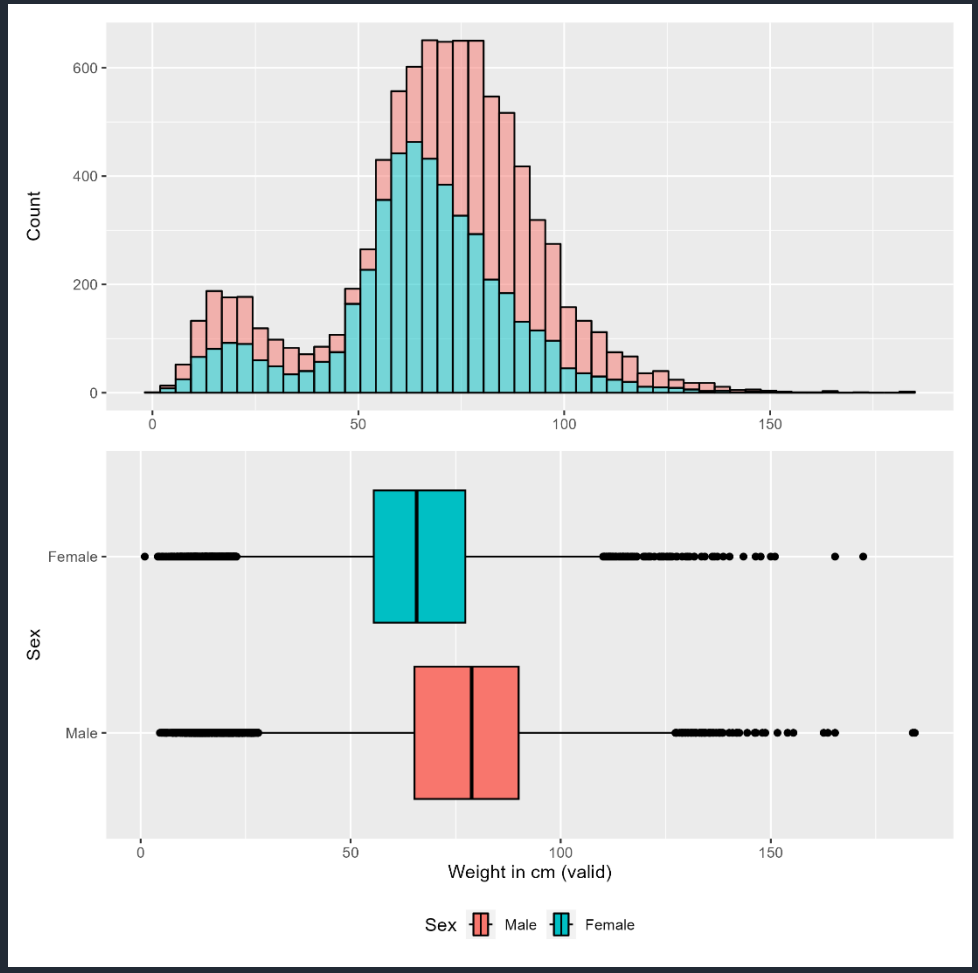
```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  data$weight
D = 0.99986, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Non-parametric Wilcoxon independent samples test

```
        Wilcoxon rank sum test with continuity correction

data:  weight by sex
W = 12449400, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```



# Inferential statistics

Q11: What is the correlation between whether a person drinks nowadays, total household income, age at last birthday, and gender?

| | Current drinker (0 = No, 1 = Yes) | Total household income | Age (at last birthday) | Sex (1 = Male, 2 = Female) |
|---|---|---|---|---|
| Current drinker (0 = No, 1 = Yes) | N/A | r=-0.073 P-value<0.001 | r=-0.069 P-value<0.001 | r=-0.116 P-value<0.001 |
| Total household income | | N/A | r=0.050 P-value<0.001 | r=0.005 P-value=0.63 |
| Age (at last birthday) | | | N/A | r=0.033 P-value<0.001 |
| Sex (1 = Male, 2 = Female) | | | | N/A |

# Discussion

- Main finding (sex and drinking):
  - Men more likely to drink than women (84% vs 74.4%)

- Health Survey for England 2022:[1]
  - Drank in last year: 84% (men) vs 78% (women)
  - Drink at least weekly: 55% (men) vs 42% (women)
  - Mean number of weekly units: 17.6 (men) vs 9 (women)

[1]NHS England (2022) *Health Survey for England*. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england.

# Conclusion

- Men were more likely to drink than women in England in 2011
- Similar findings in 2022, but with a worsening for women
- Further educational efforts on alcohol-related harm should be implemented to curb alcohol consumption
- Campaigns tailored to men may be important in bringing alcohol consumption among men closer to women
- Campaigns designed for women are also needed to reverse secular trend of increase

# Appendix

Analysis code and outputs at:

https://github.com/gpessoaamorim/artificial_intelligence_pgdip/tree/master/Assignments/Module%202/Statistical%20analysis%20presentation

# References


- NHS England (2022) *Health Survey for England*. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england.
- NHS (2022) ‘Risks - Alcohol misuse’. Available from: https://www.nhs.uk/conditions/alcohol-misuse/risks/.