

Study Of Applicants Applying for Masters and PhD programs at UofR

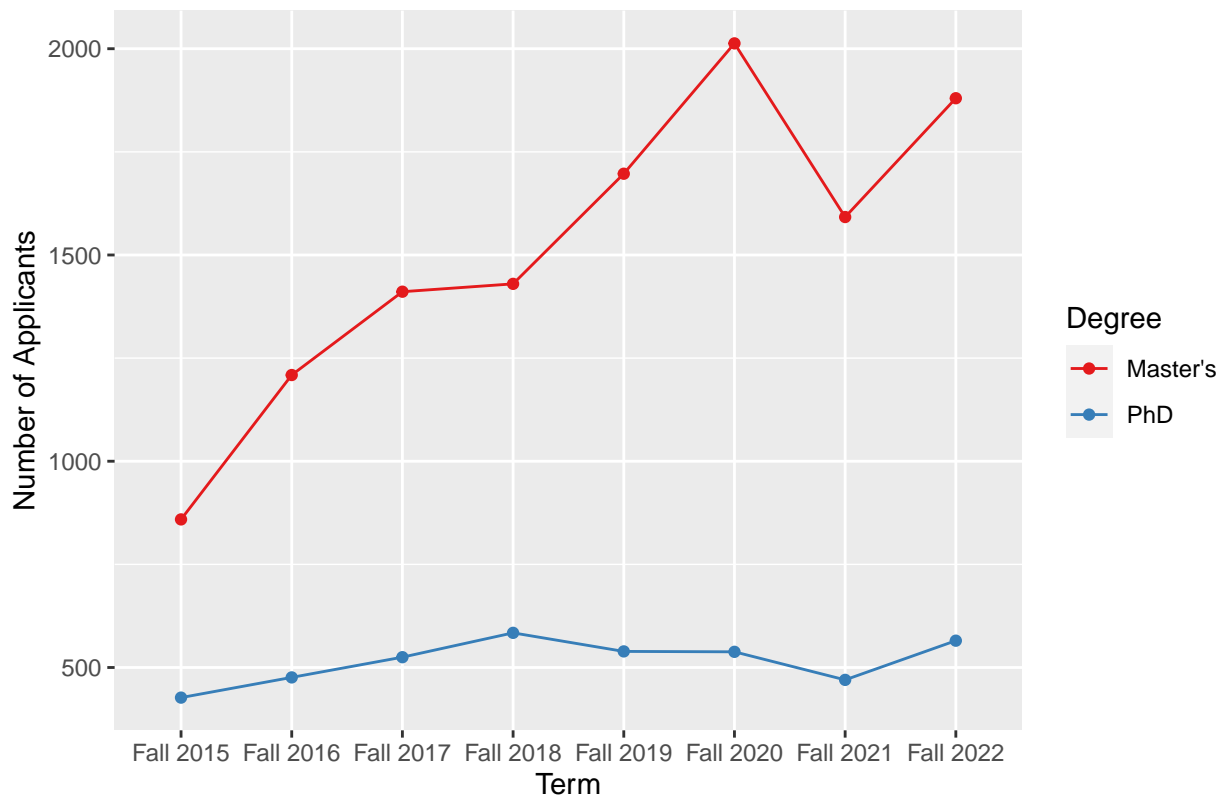
Anik De, Babli Dey, Sharon Gilbert, Veronica Mata

2023-12-14

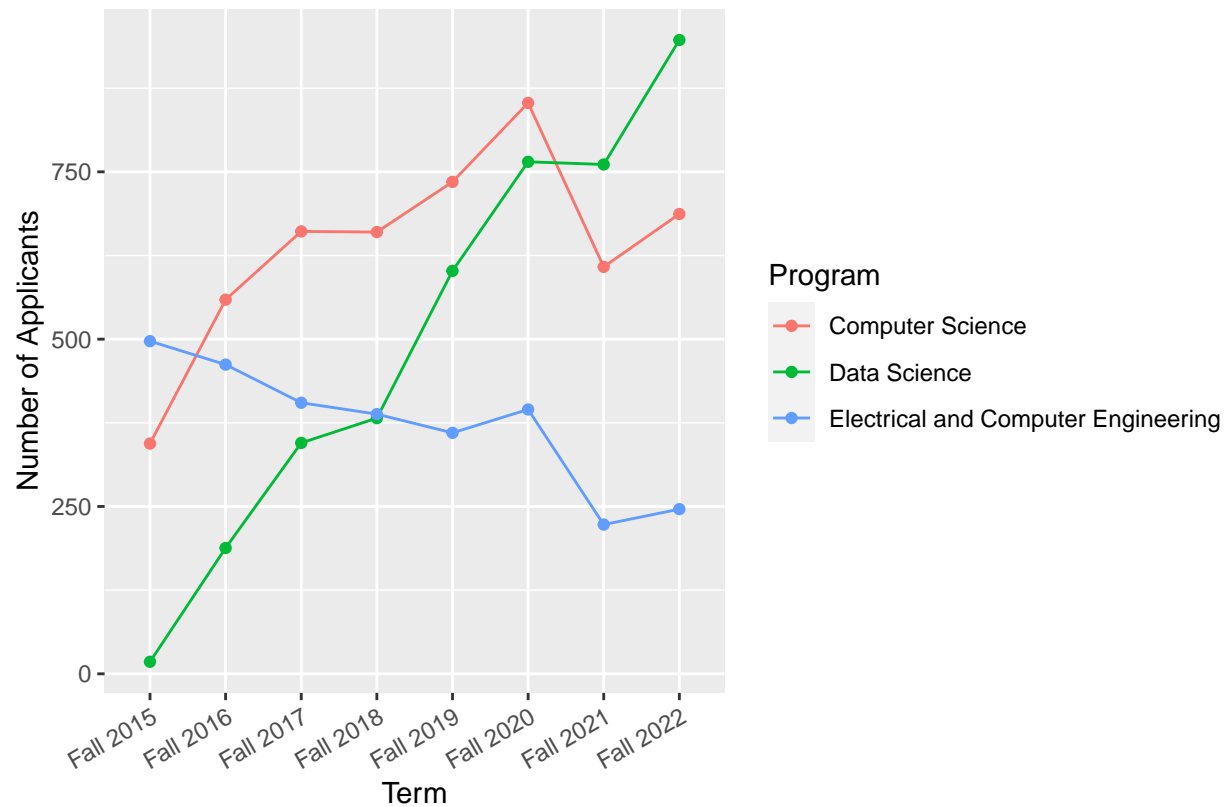
Visualizations of EDA (will include unknown data) - Masters v/s PhD:

1. What is the annual total no. of applicants, which term recorded the highest and lowest number of applicants?

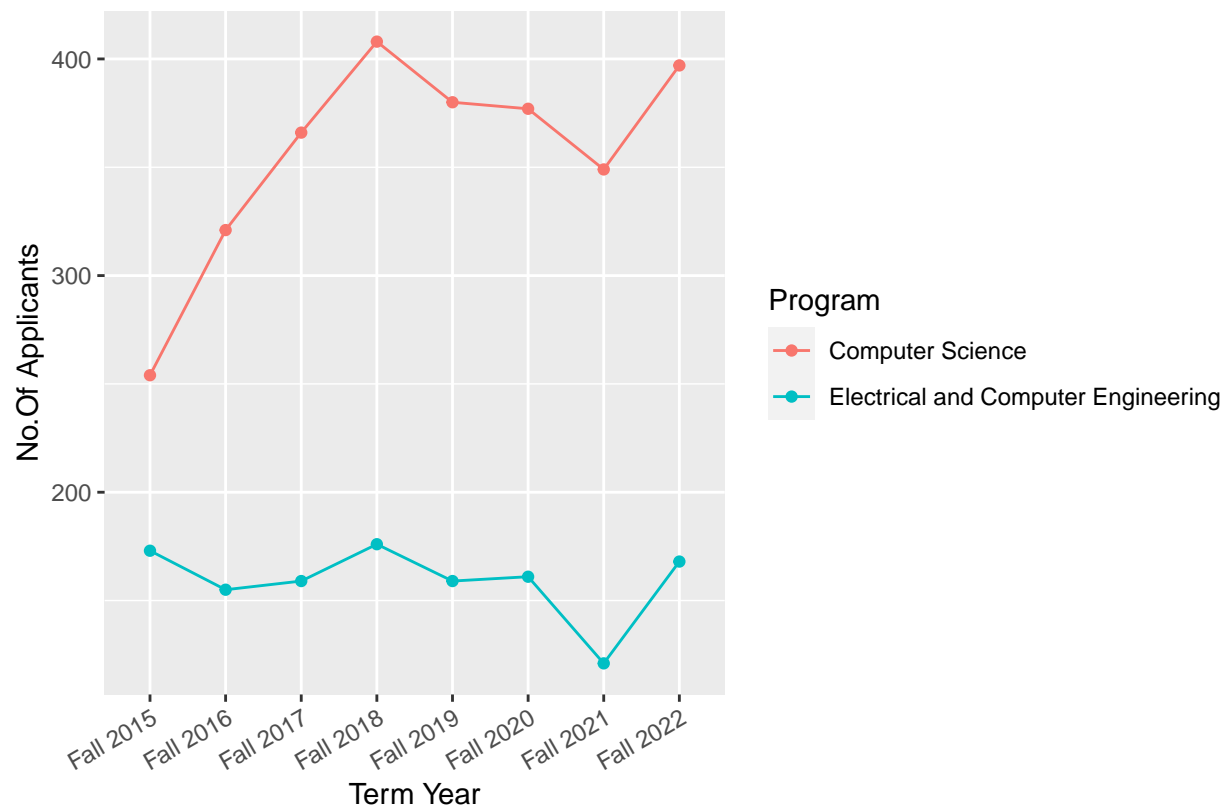
U of R Student Application – Number of Applicants by Program (2015 – 2022)



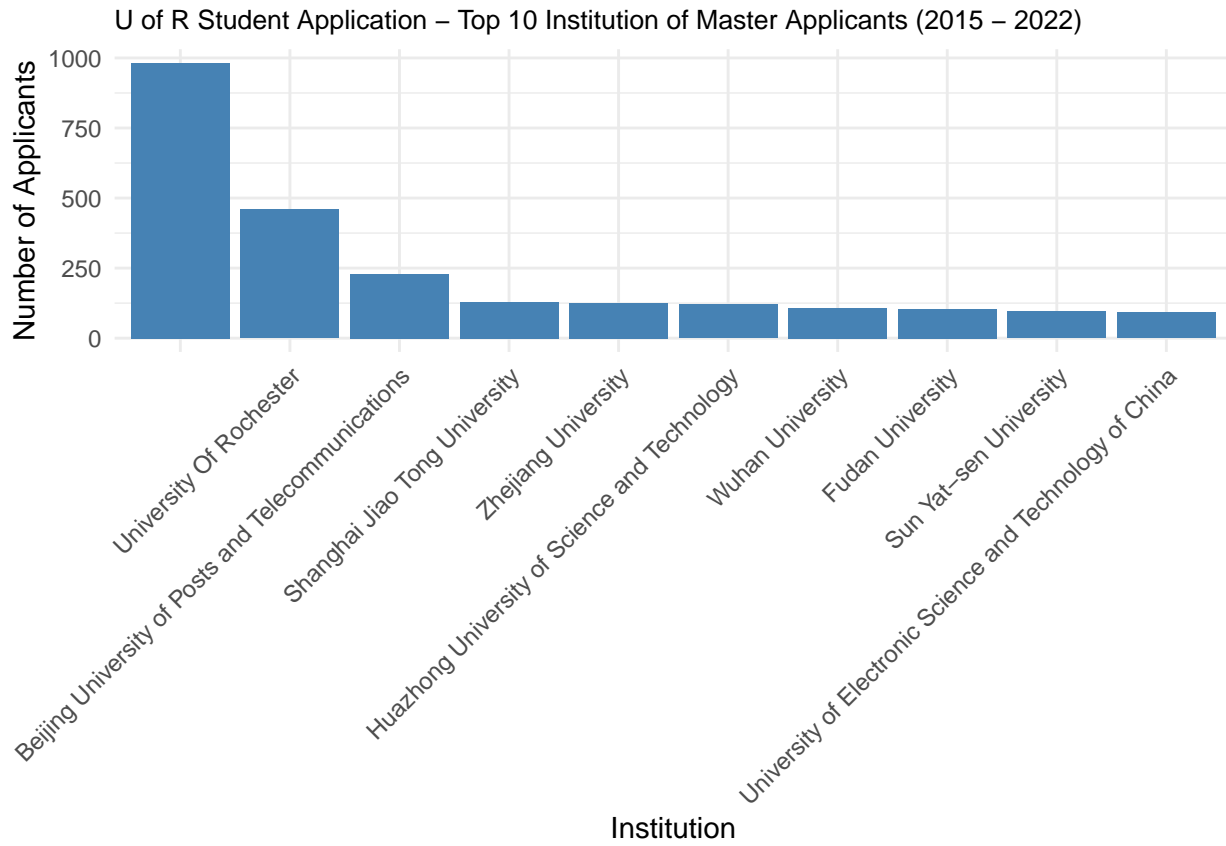
Student Application – Number of Master's Sub-Program Applicants by Program (2015 – 2022)

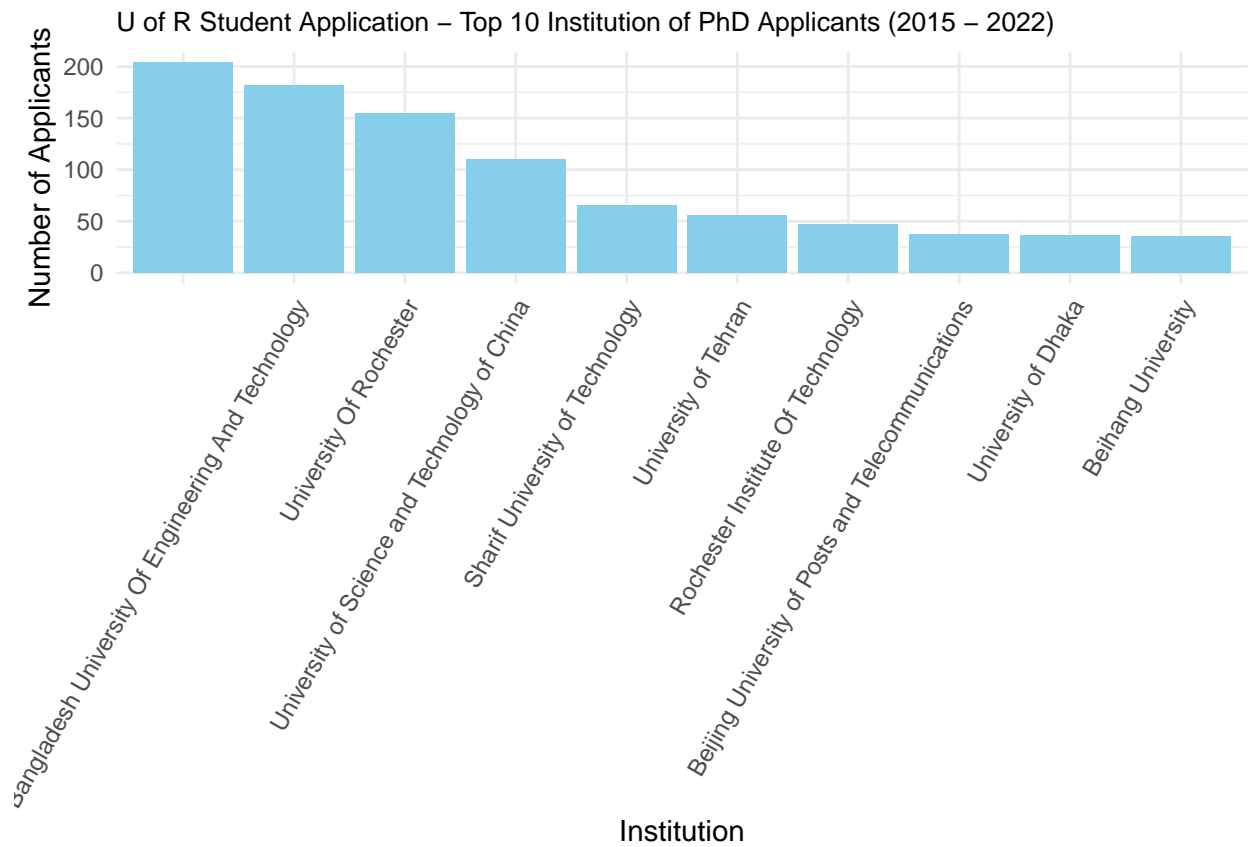


nt Application – Total No. of PhD Sub Program Applicants Every Fall Term By Program (2015 – 2022)



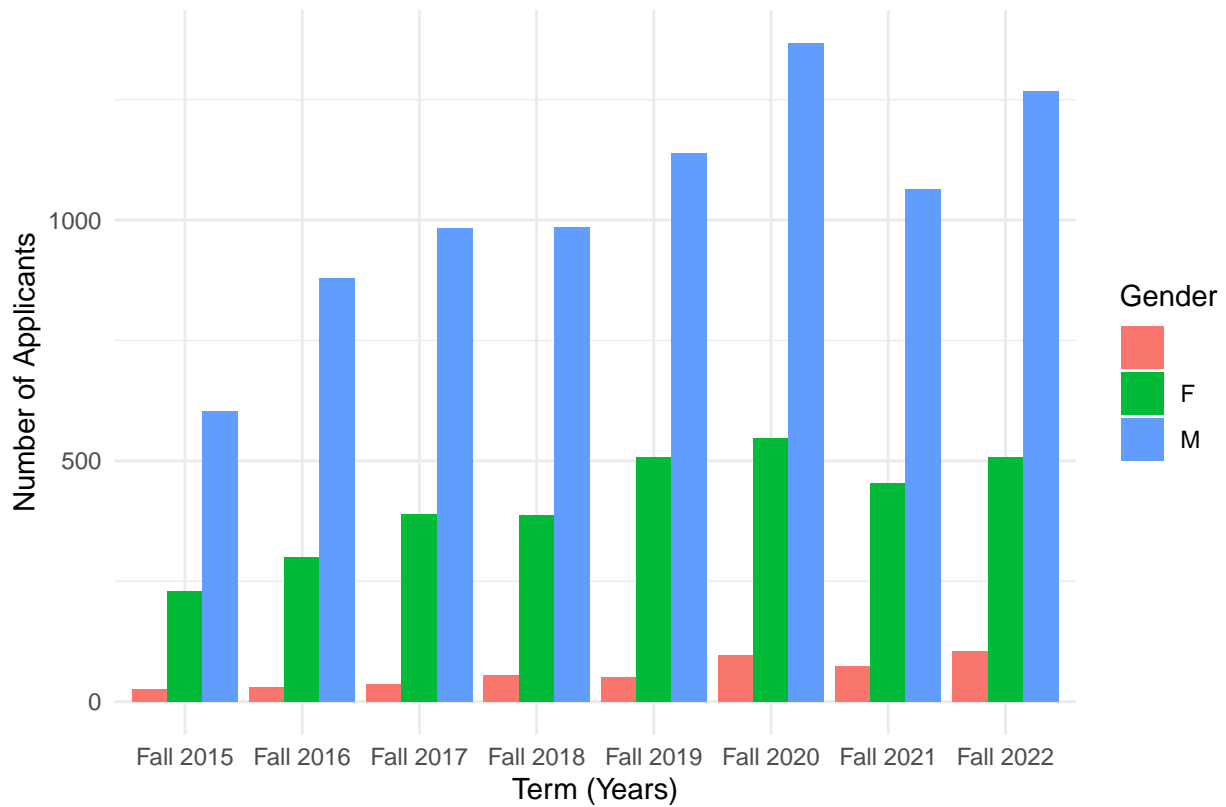
2. Identify the top 10 undergraduate institutes from which applicants are applying.



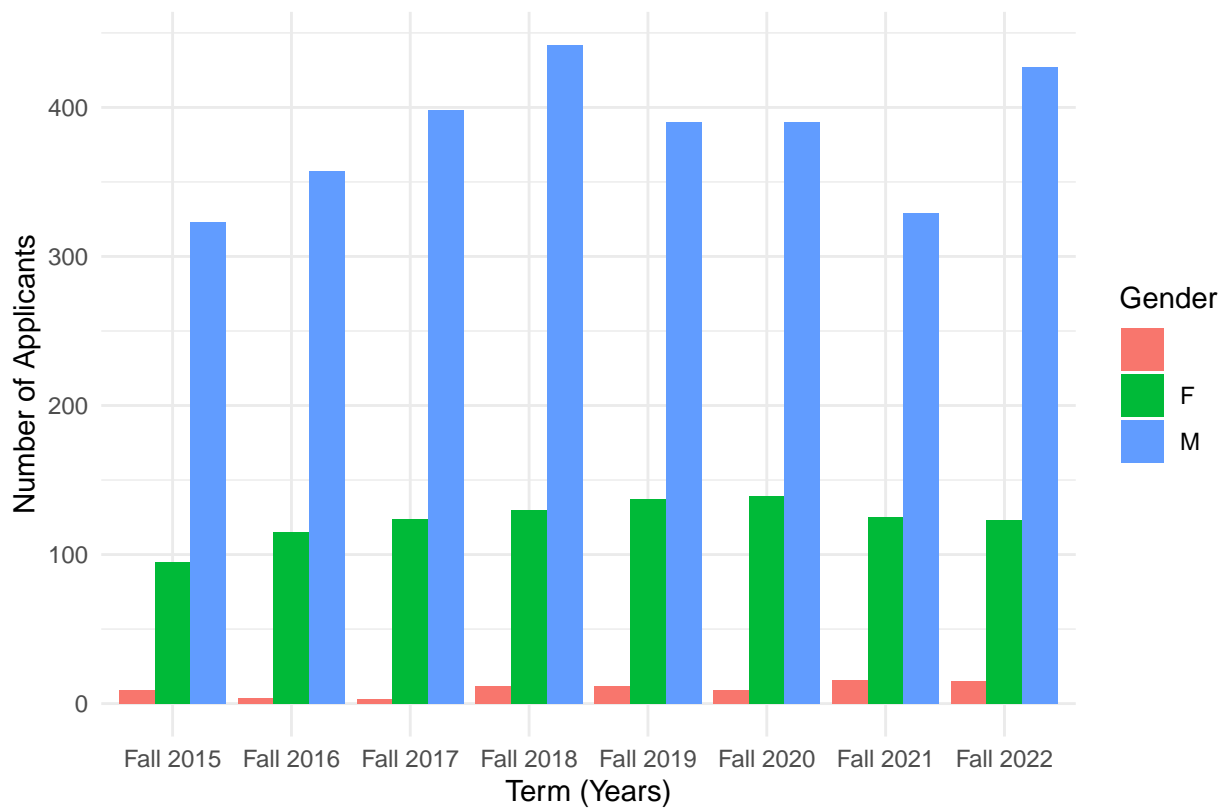


3.How do the numbers of applicants vary by gender across different terms?

U of R Student Application – No. Of Applicants for Master's by Gender and Term (2015 – 2022)

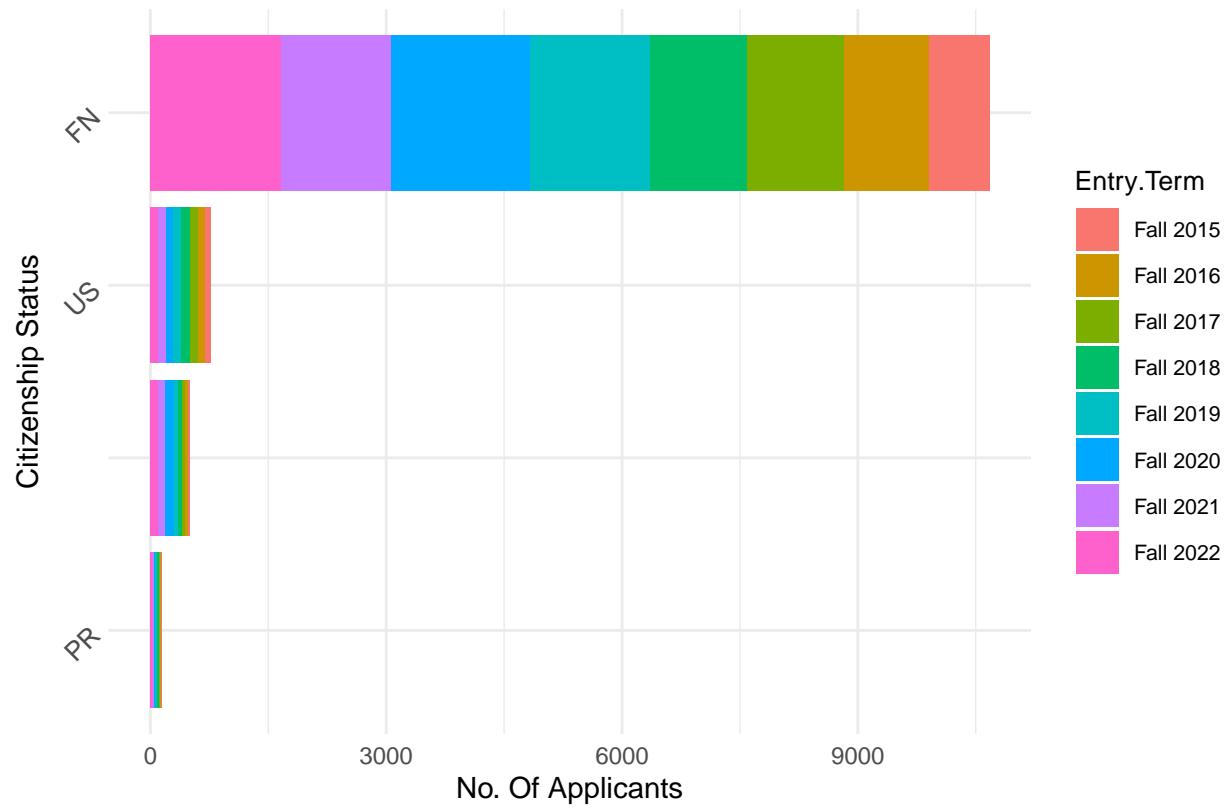


U of R Student Application – No. Of Applicants for PhD by Gender and Term (2015 – 2022)

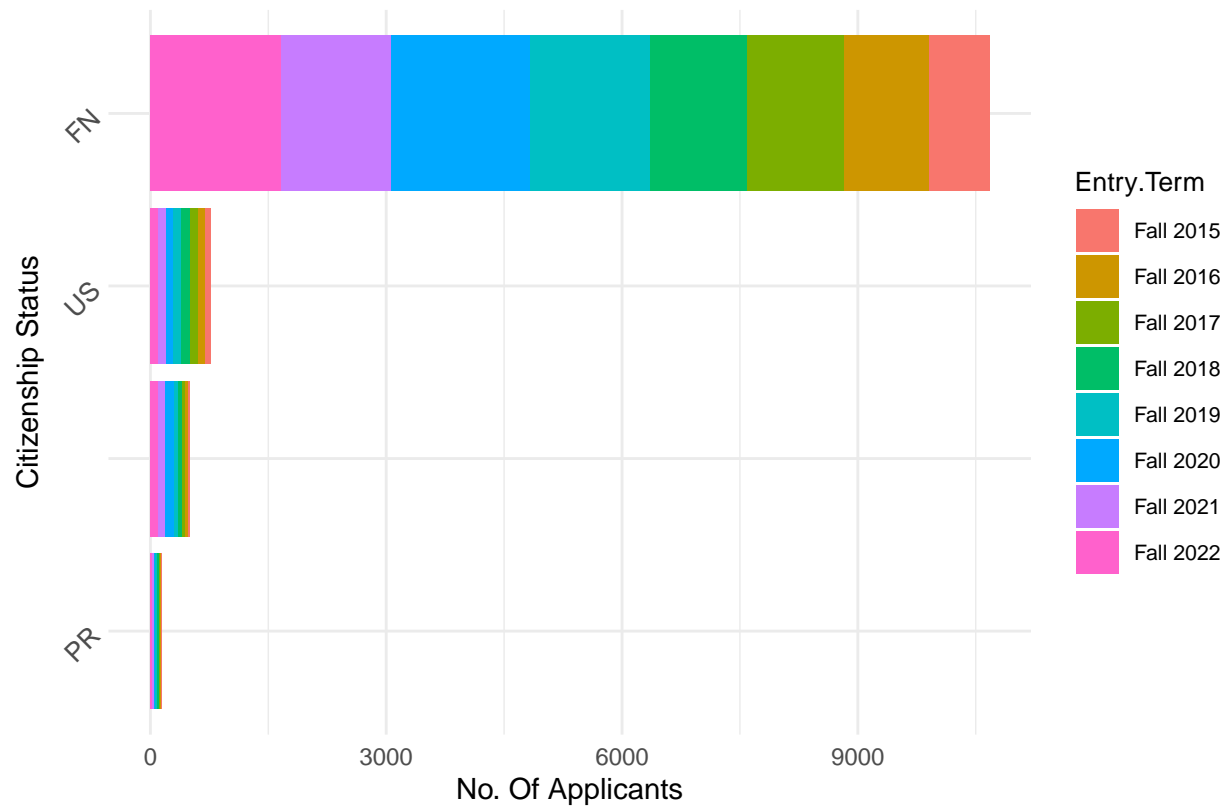


4.What are the top 10 countries from which applicants have applied to the university?

of R Student Application – No. Of Master's Applicants based on Citizenship By Term (2015 – 2022)

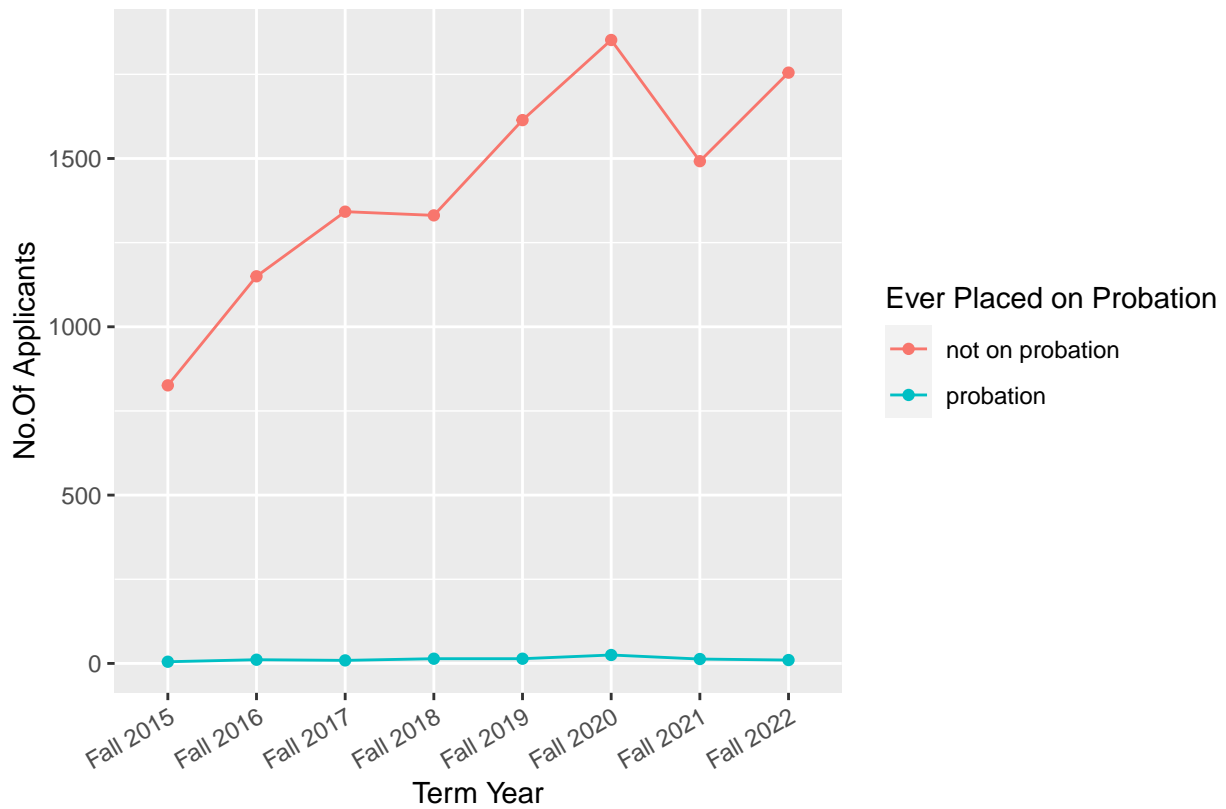


U of R Student Application – No. Of PhD Applicants based on Citizenship By Term (2015 – 2022)

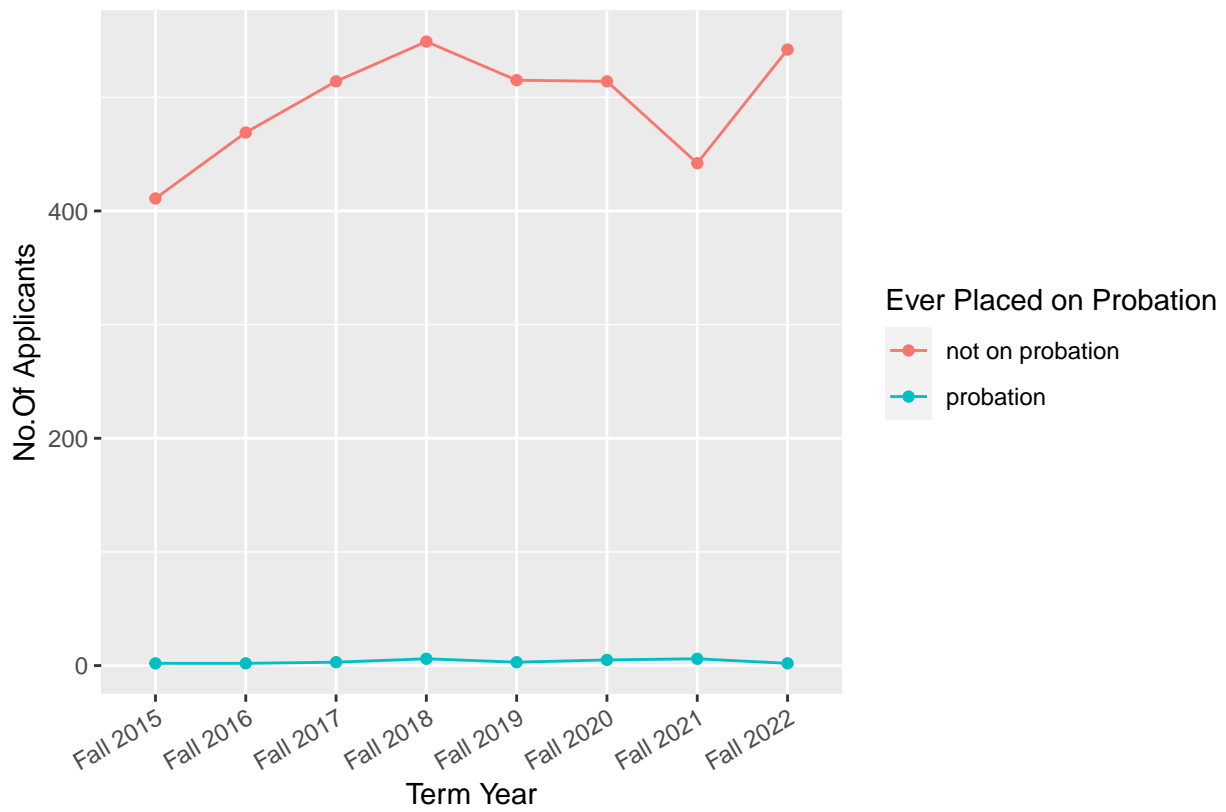


5. Determine the total number of applicants who were placed on probation.

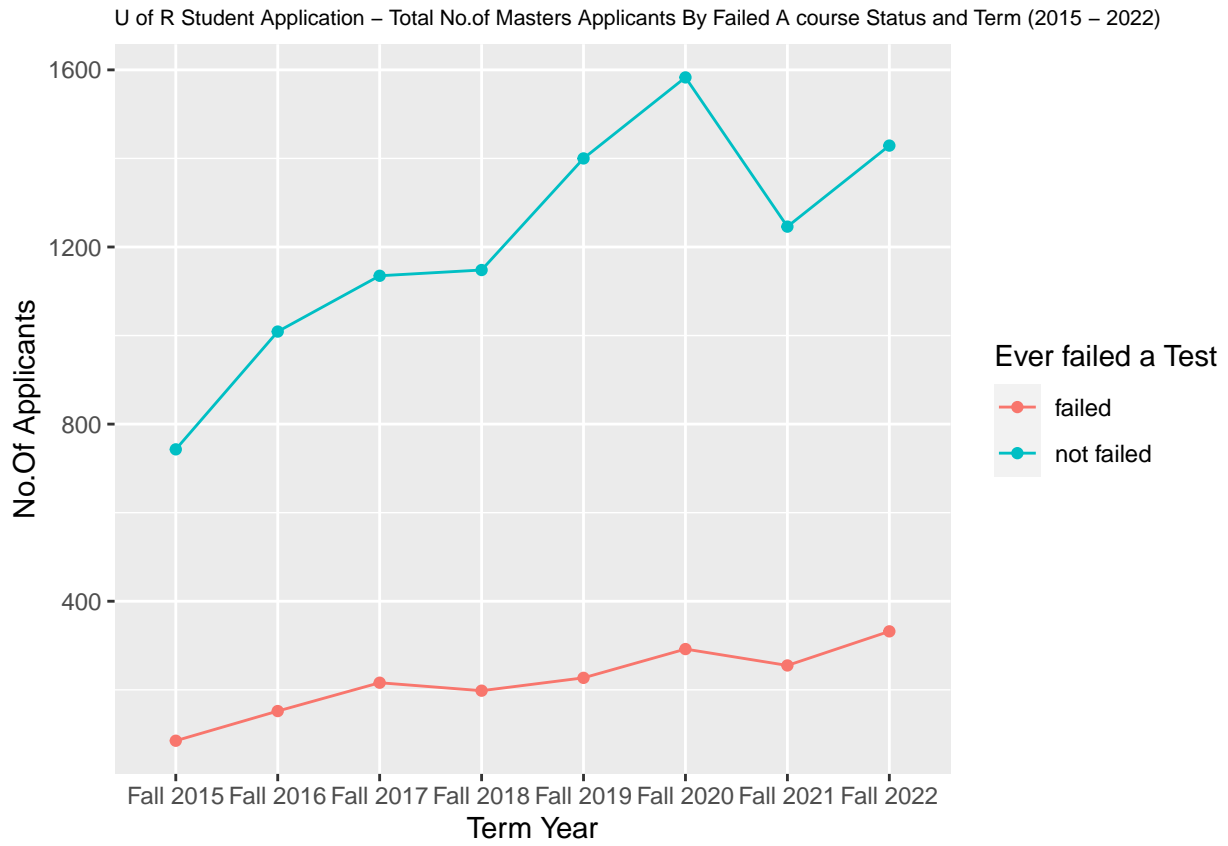
U of R Student Application – Total No.of Masters Applicants By Probation Status and Term (2015 – 2022)

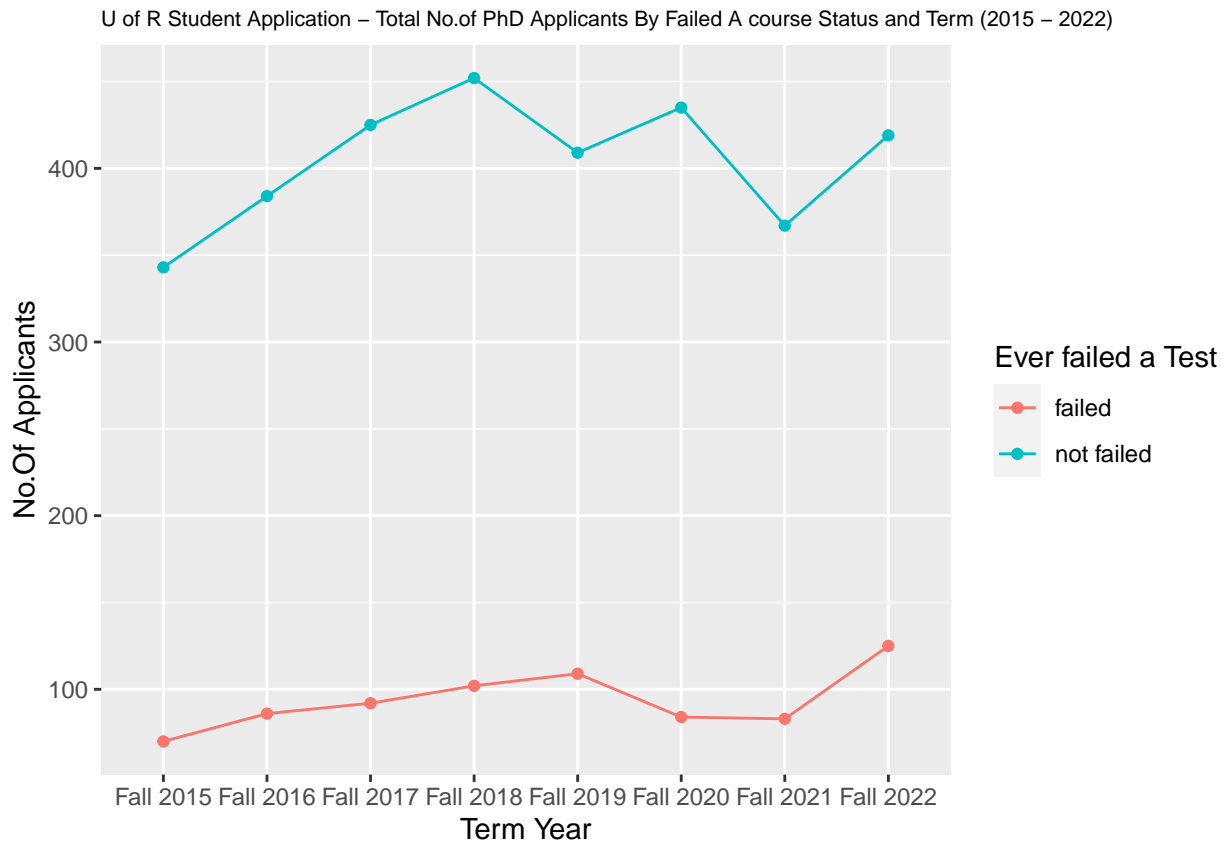


U of R Student Application – Total No.of PhD Applicants By Probation Status and Term (2015 – 2022)



6. Determine the total number of applicants who failed a course.





Data Cleaning: (code in rmd file)

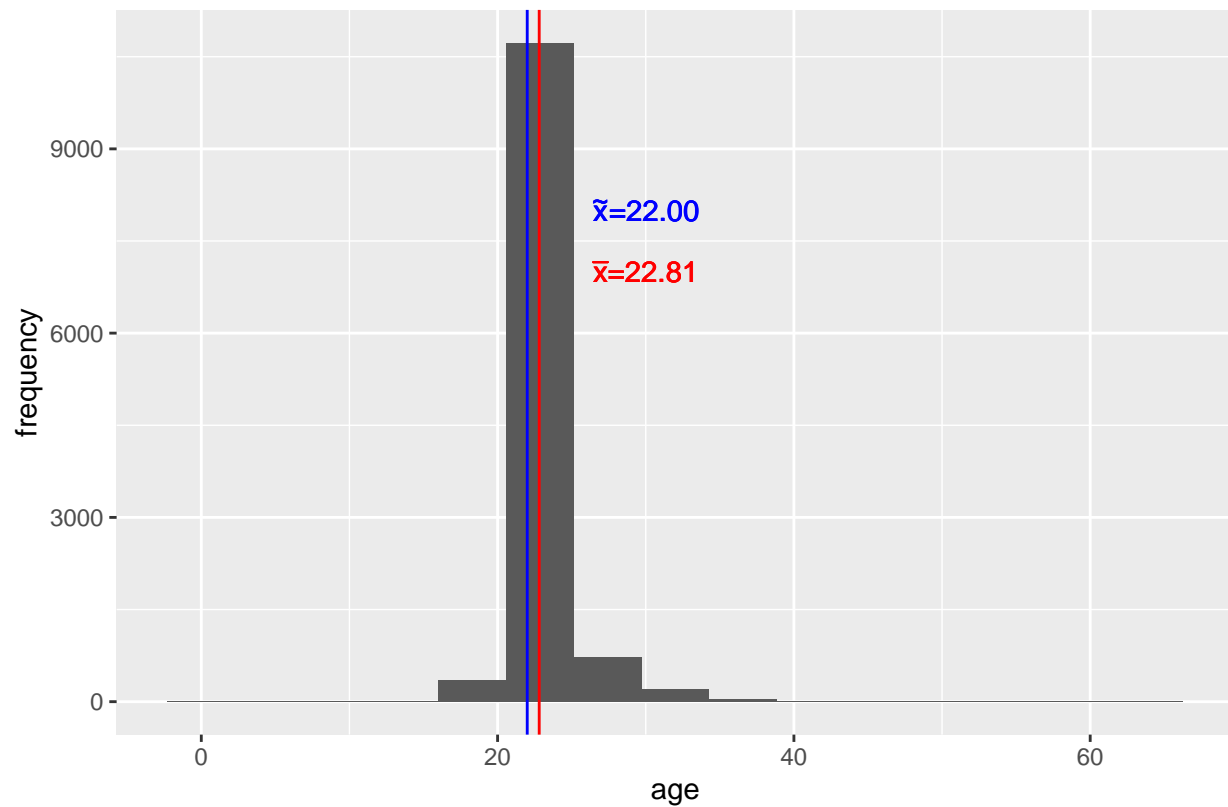
1. Apply a filter to the dataset based on the 24 specified dimensions.
2. Change column names to lowercase and use underscores or simpler terms for better interpretability.
3. Segment the dataset into two parts: one for master's programs and another for Ph.D. programs.
4. For empty values, categorize them as "Unknown" for categorical data. For age, fill in with the average age corresponding to the respective program.

Descriptive Analysis: Master's and PhD

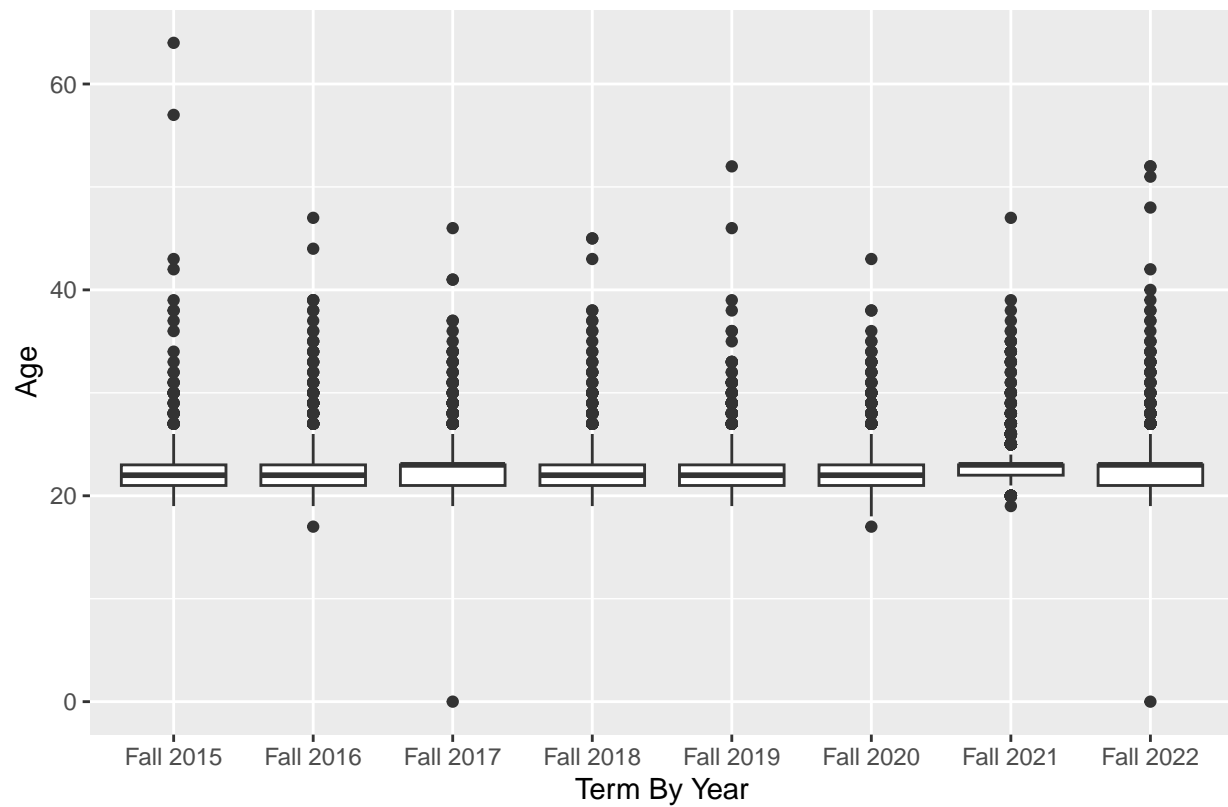
1. Calculate the mean, variance, and standard deviation of age, and visualize the distribution using histograms and box plots.

```
## The Mean age of Masters applicant is: 22.80763
##
## The Variance in age of Masters applicant is: 6.116835
##
## The Standard Deviation in age of Masters applicant is: 2.473224
```

U of R Student Application – Histogram of Master's Applicant Age (2015 – 2022)



U of R Student Application – Box Plot of Master's Applicant Age By Term (2015 – 2022)



```
## The Mean age of PhD applicant is: 24.92992
```

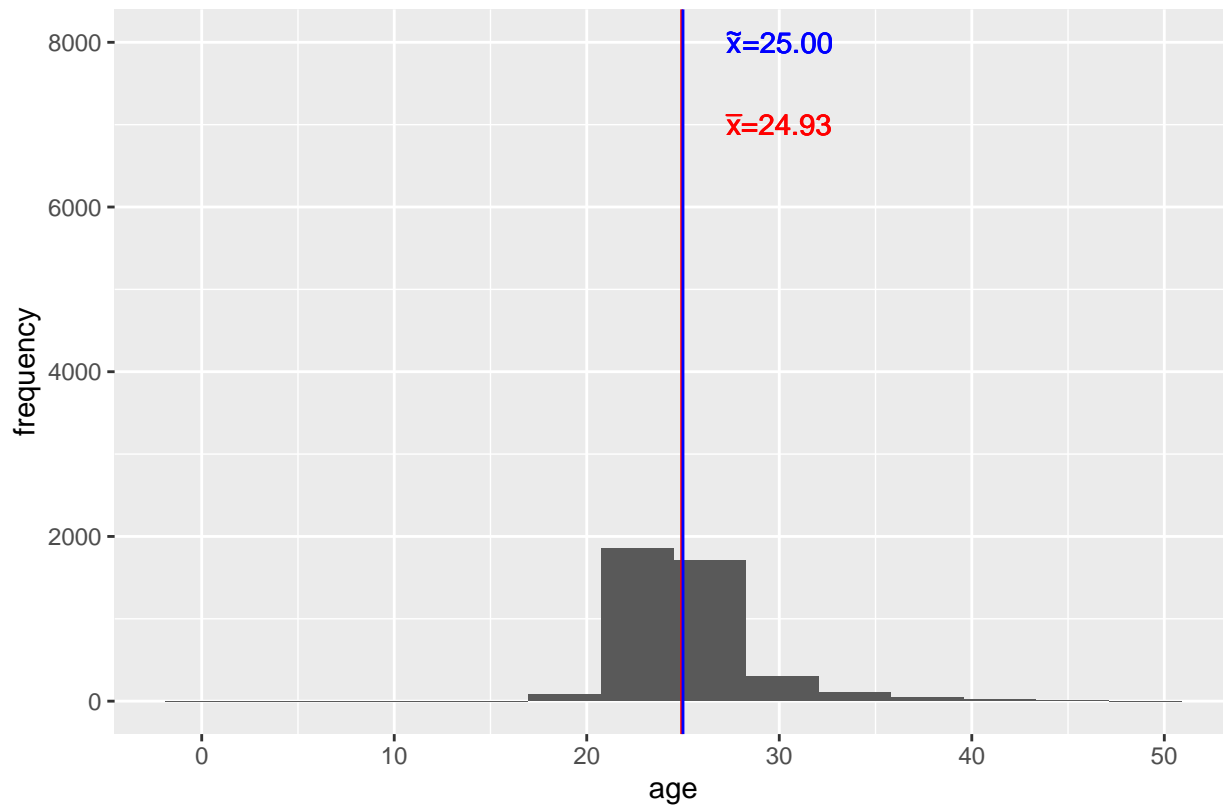
```
##
```

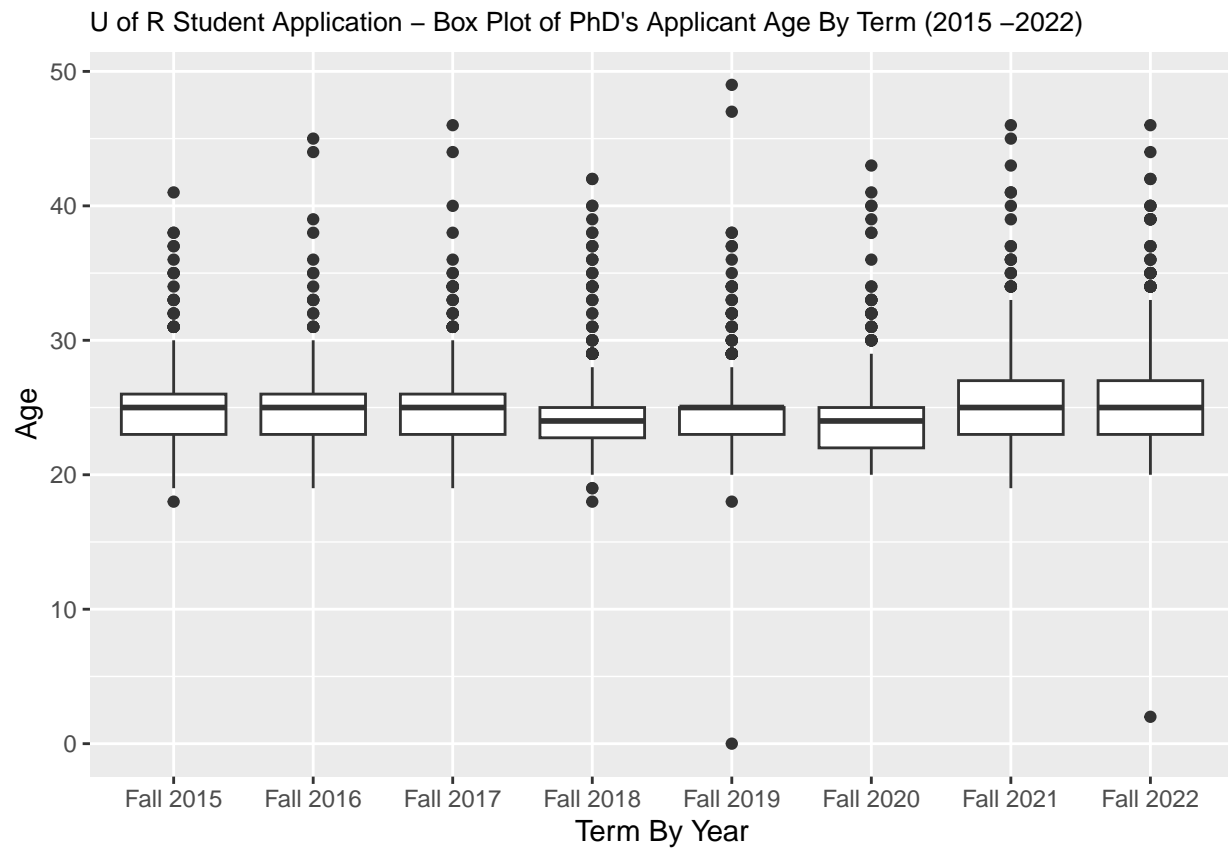
```
## The Variance in age of PhD applicant is: 12.33731
```

```
##
```

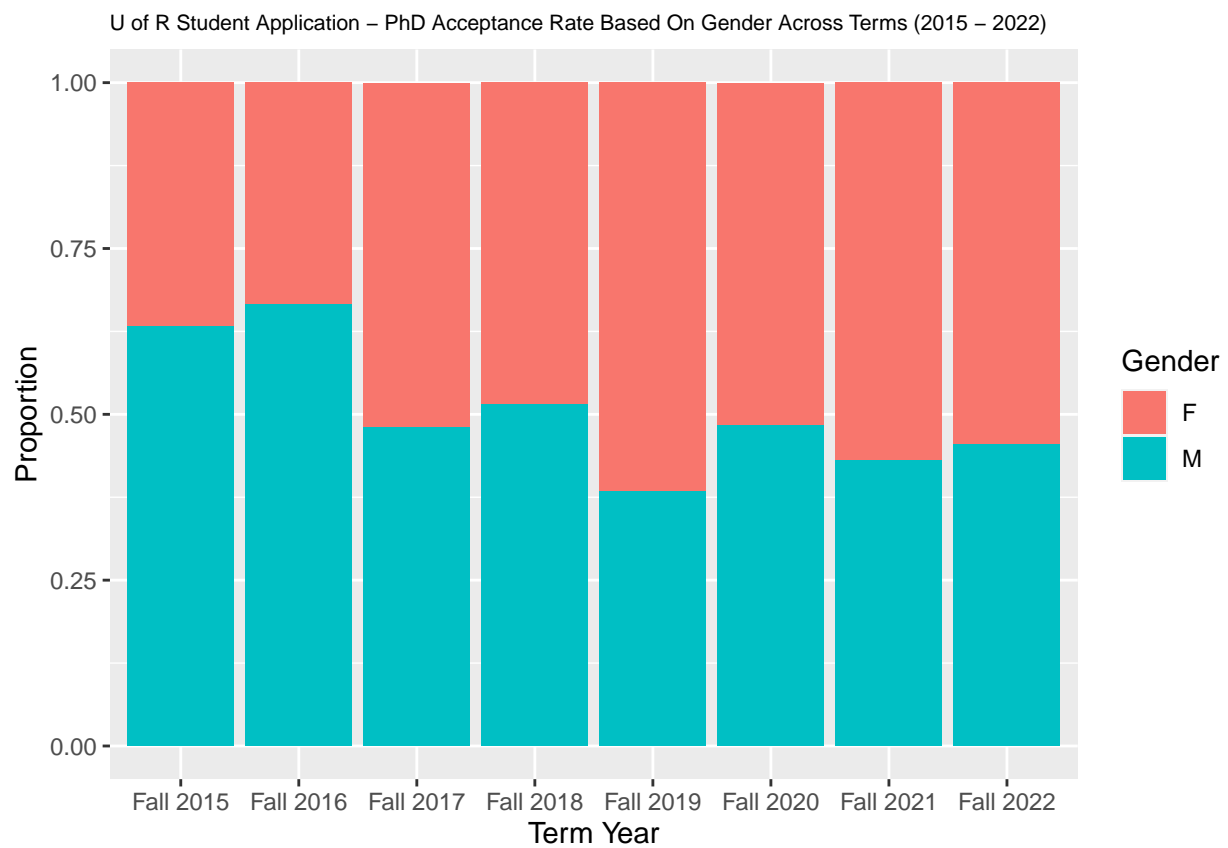
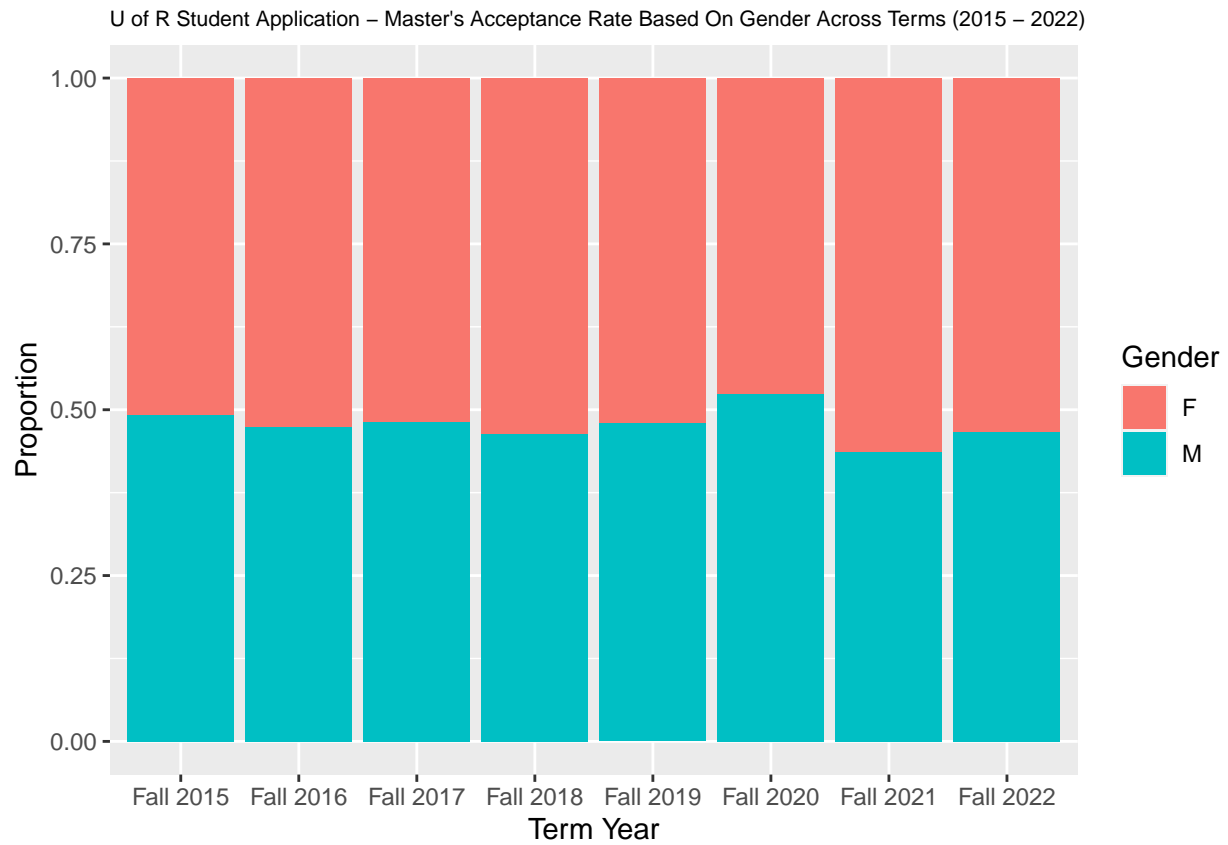
```
## The Standard Deviation in age of PhD applicant is: 3.512451
```

U of R Student Application – Histogram of PhD Applicant's Age (2015 – 2022)

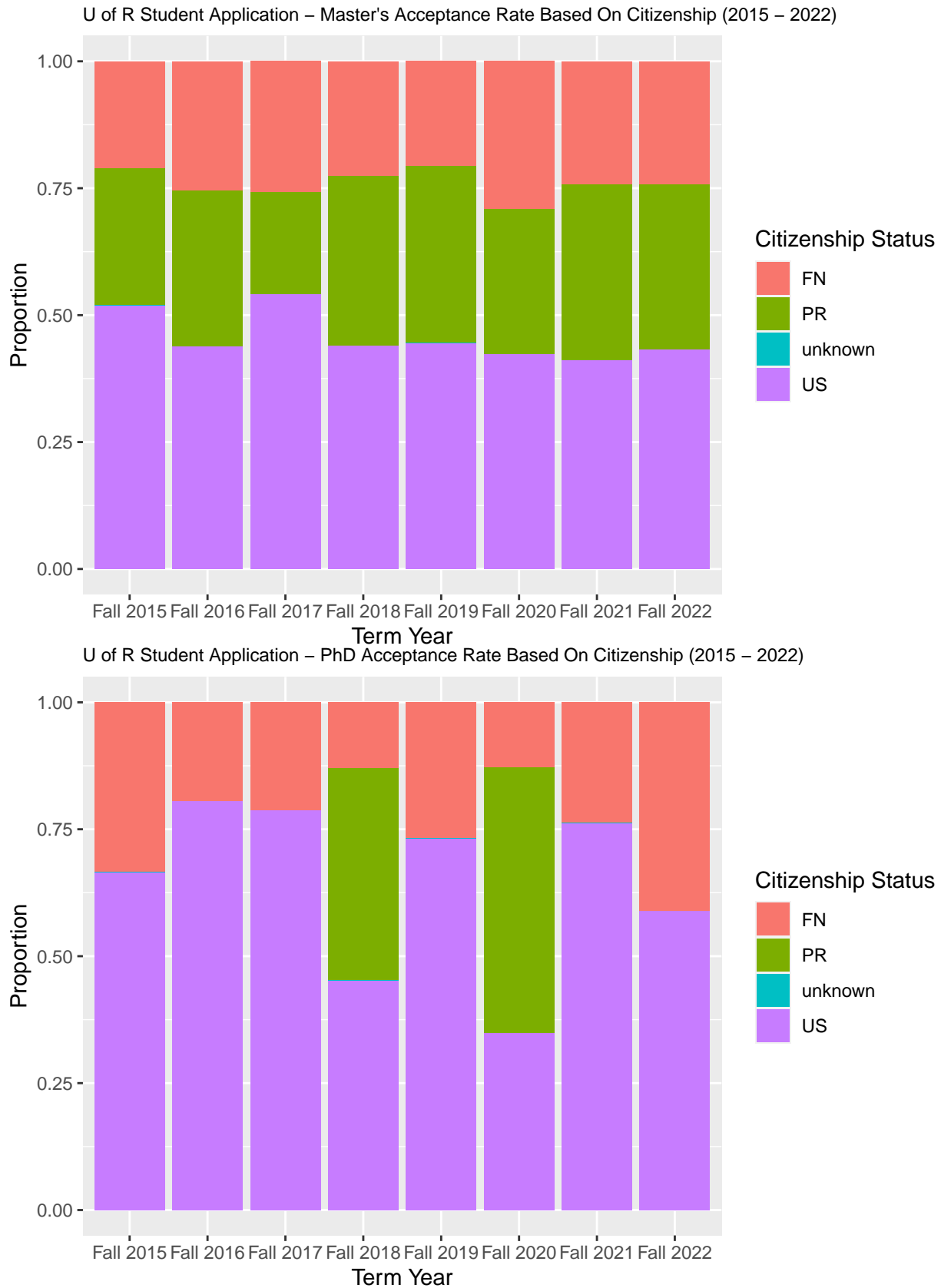




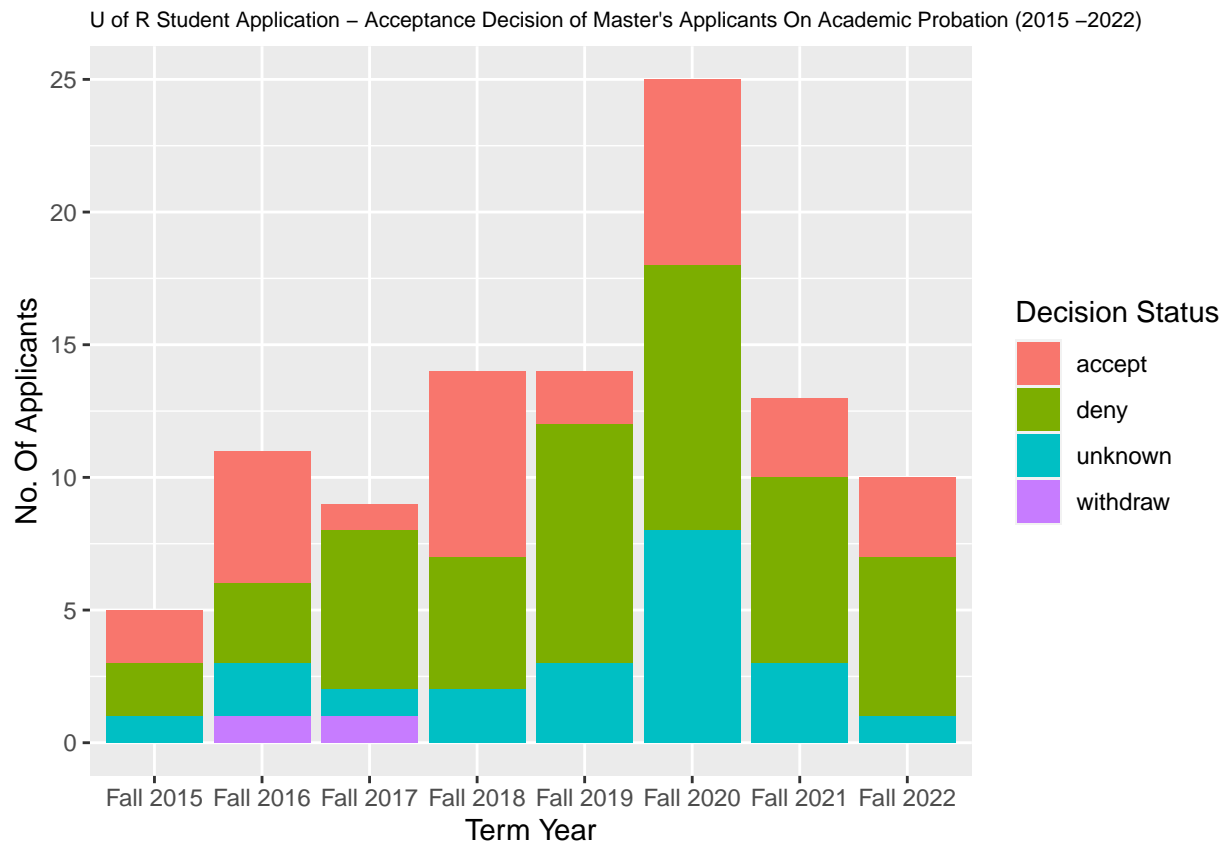
2.Examine the acceptance rate based on gender across different terms.

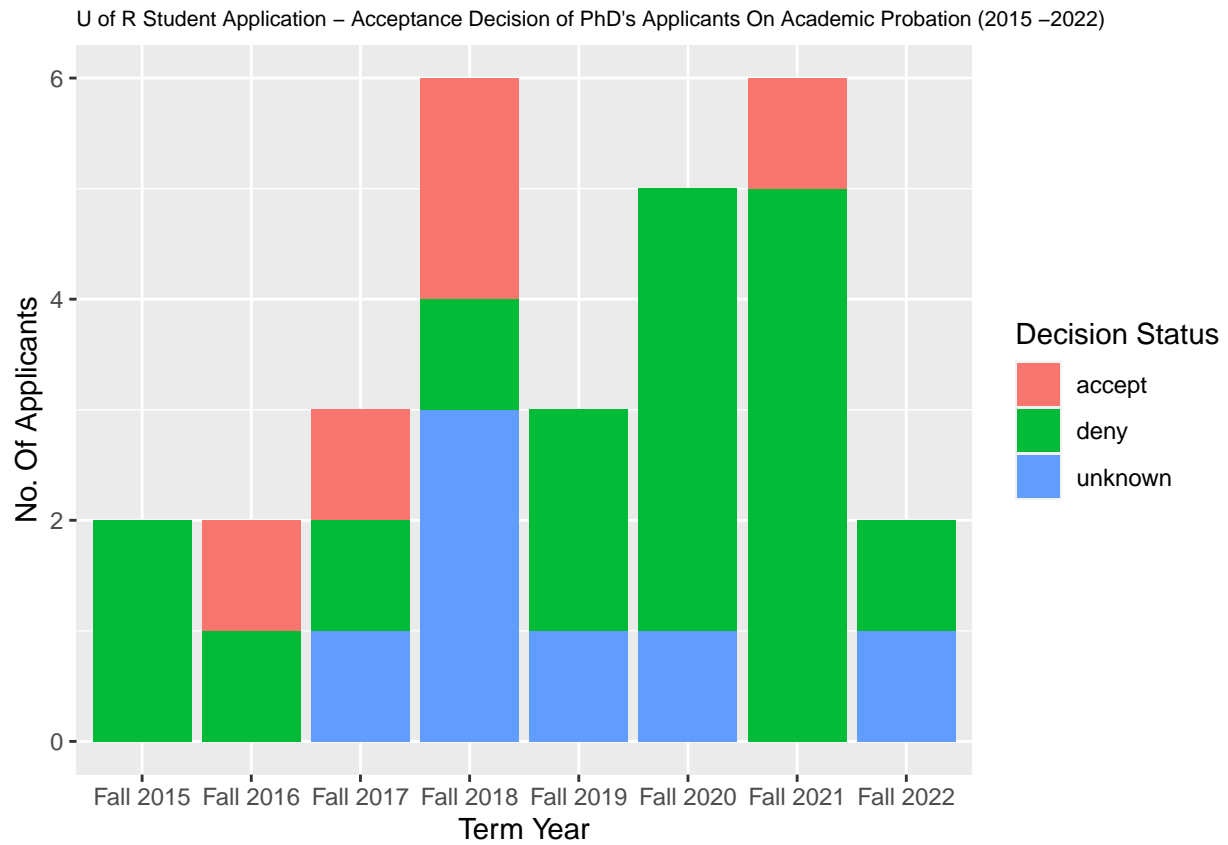


3.Explore the acceptance rate by country in relation to different terms.



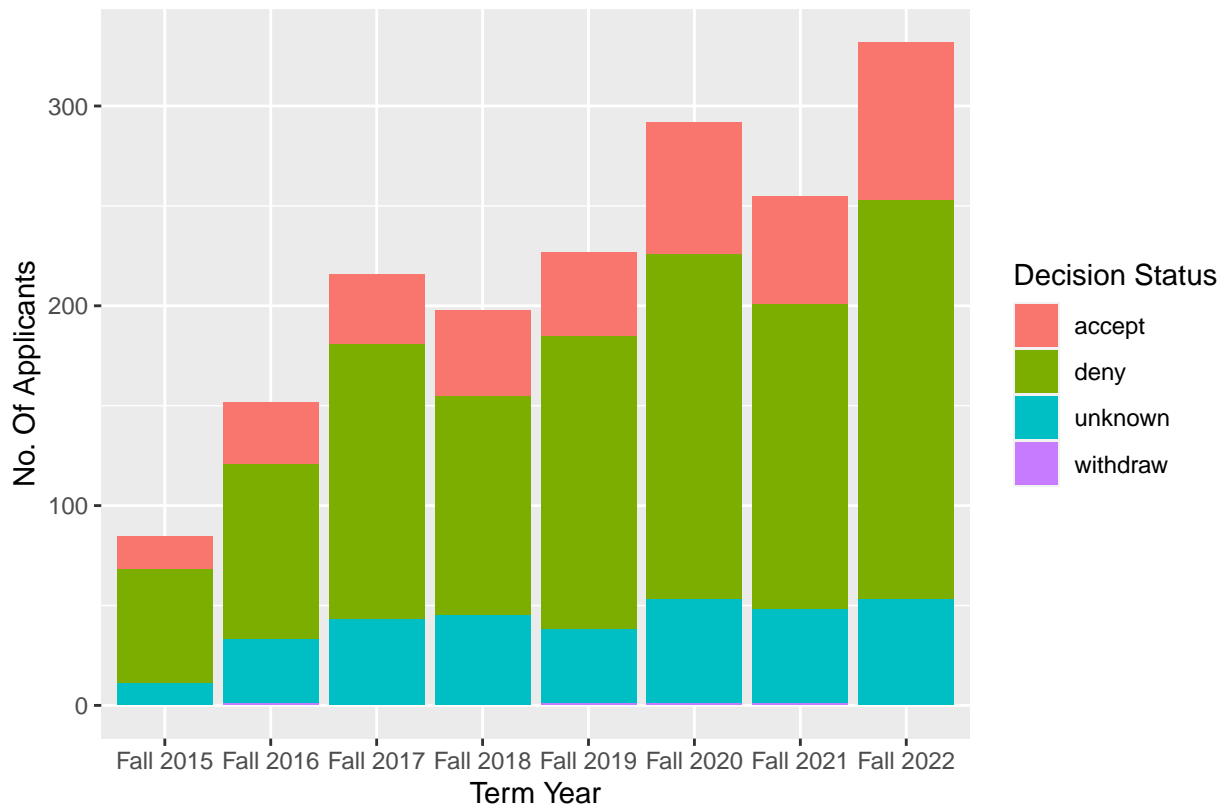
4. Investigate the acceptance decision for applicants who were placed on probation.



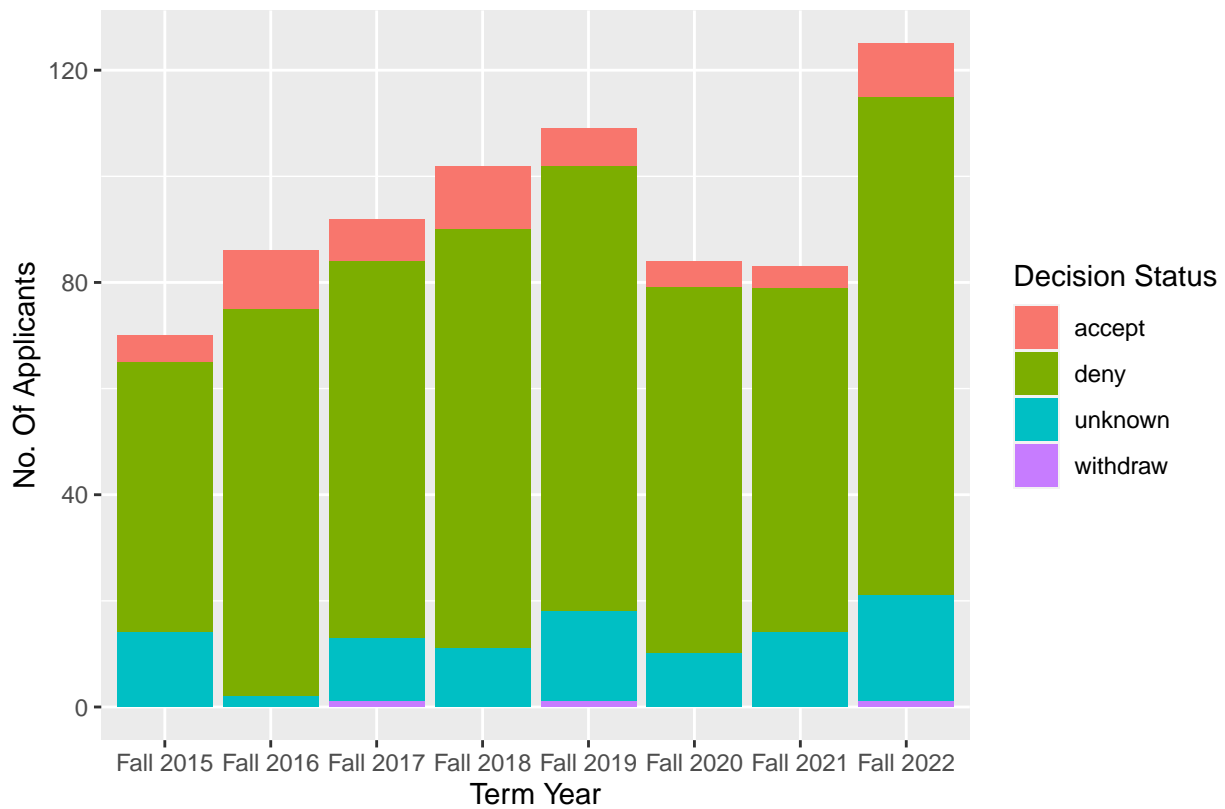


5. Investigate the acceptance decision for applicants who failed a course.

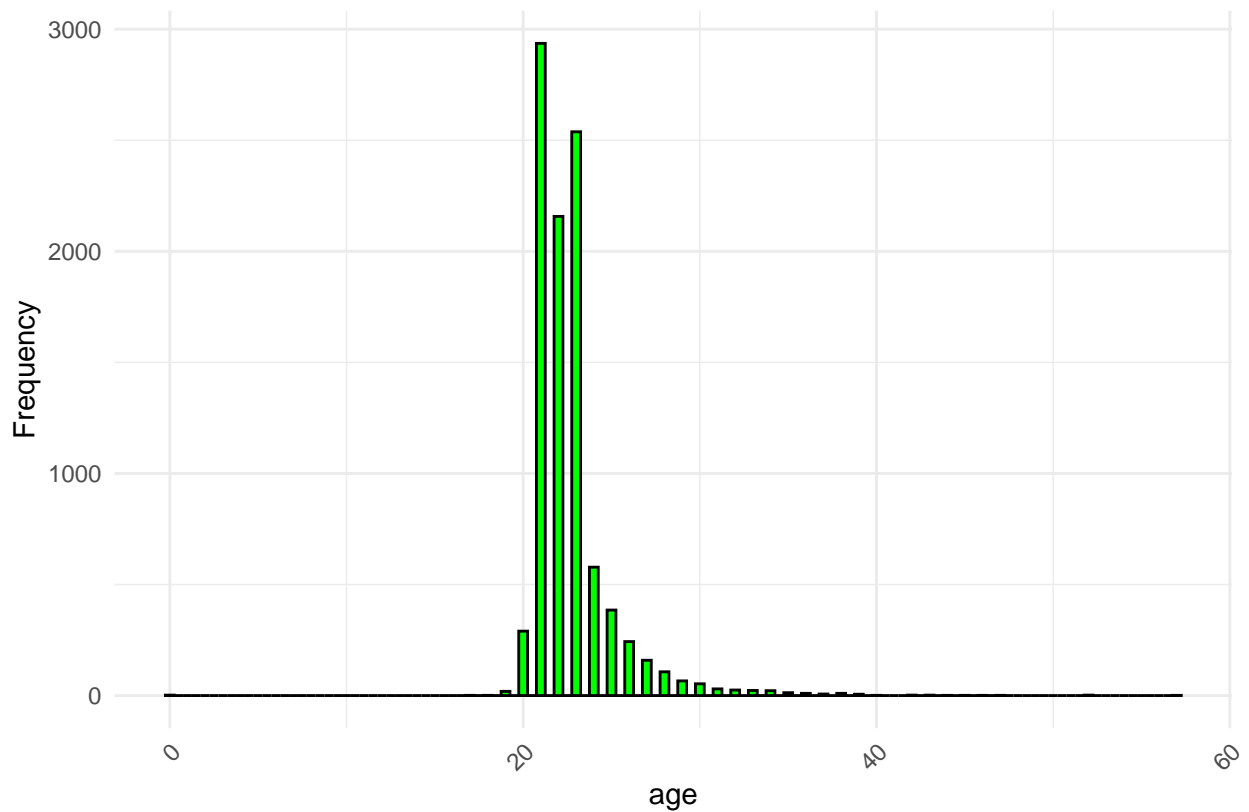
U of R Student Application – Acceptance Decision of Master's Applicants who failed a course (2015 – 2022)



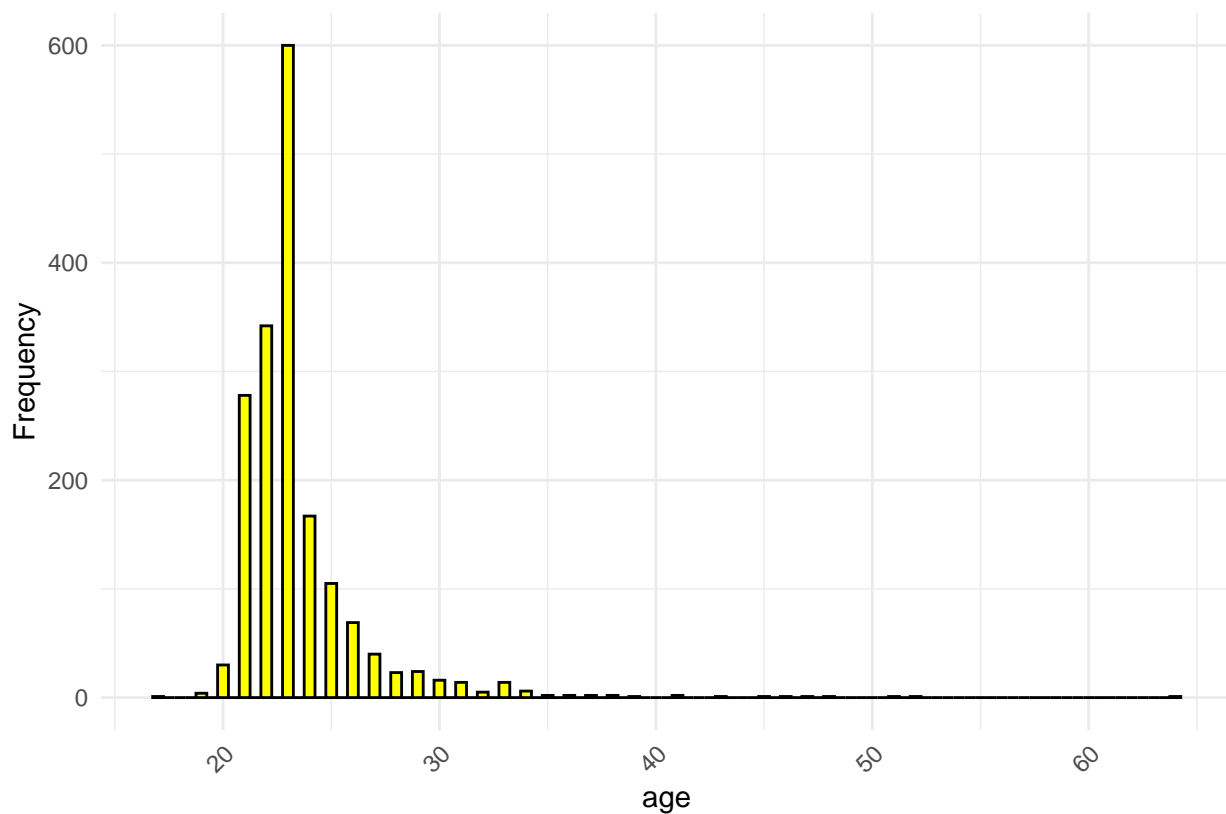
U of R Student Application – Acceptance Decision of Master's Applicants who failed a course (2015 – 2022)

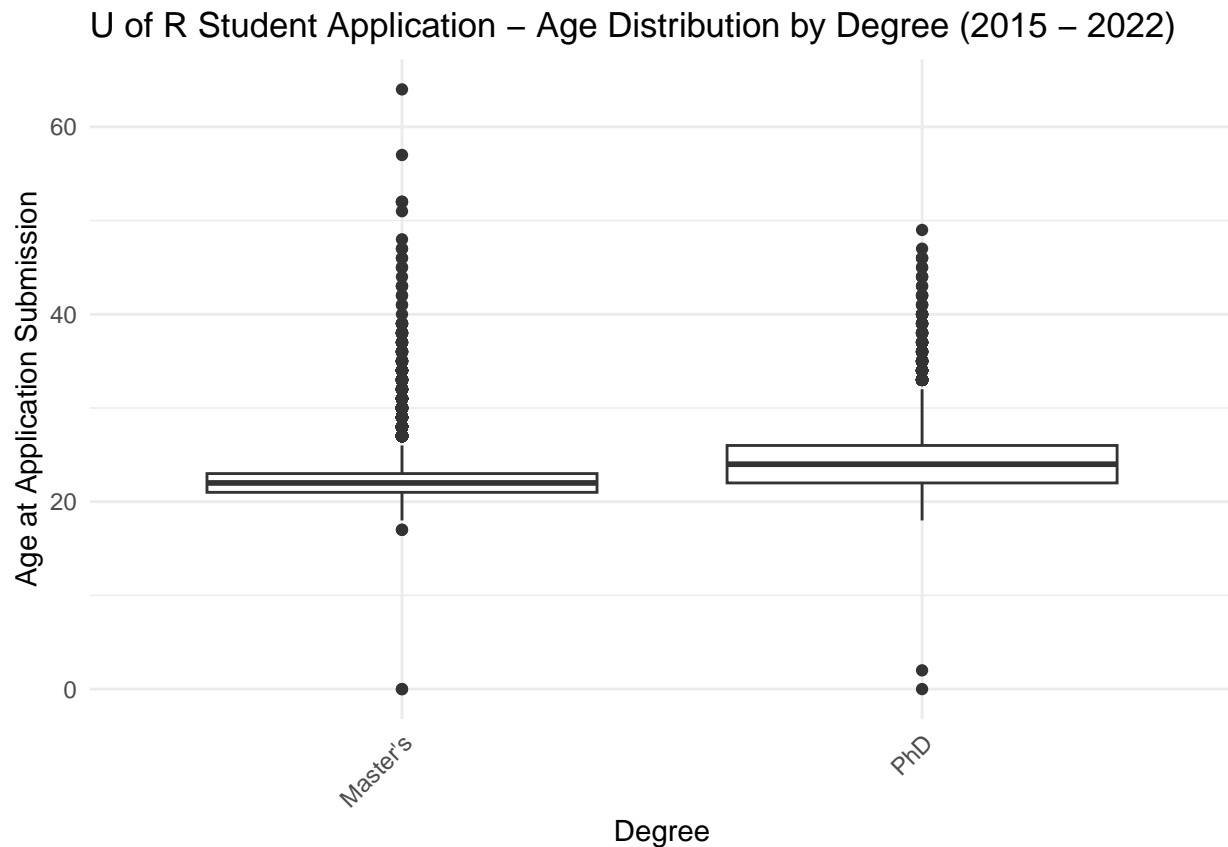


6. Investigate the age distribution between applicants who failed a course v/s applicants who did not fail a course.
 U of R Student Application – Distribution of age of Master's Applicants Never Failed a Course (2015 – 2022)



U of R Student Application – Distribution of age of Master's Applicants Failed a Course (2015 – 2022)





Inferential Analysis:

1. Is there a notable distinction in age variance between applicants who have experienced course failure and those who haven't, considering each degree type?

H_0 : There is no difference in the variance of ages between Master's applicants who have experienced course failure and those who haven't. (i.e., the variances are equal)

H_A : There is a significant difference in the variance of ages between Master's applicants who have experienced course failure and those who haven't.

```
## [1] 1757
## [1] 9693
##
## F test to compare two variances
##
## data:  masters_failed and masters_not_failed
## F = 1.7056, num df = 1756, denom df = 9692, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.588744 1.834474
## sample estimates:
## ratio of variances
##      1.705644
##
```

```
## F test to compare two variances
##
## data: phd_failed and phd_not_failed
## F = 1.1811, num df = 750, denom df = 3233, p-value = 0.003029
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 1.057599 1.324210
## sample estimates:
## ratio of variances
## 1.181061
```

Conclusion : Given that the p-value is much less than the significance level ($\alpha = 0.05$), we reject the null hypothesis. This suggests that there is a statistically significant difference in the variance of the ages between Master's applicants who have experienced course failure and those who haven't. (1)

2. Is there a discernible difference in the average age between master's and Ph.D. students?

H_0 : The average age of Master's students is equal to the average age of Ph.D. students. (no difference in means)

H_A : The average age of Master's students is not equal to the average age of Ph.D. students.

```
##
## Welch Two Sample t-test
##
## data: masters_data$age and phd_data$age
## t = -35.886, df = 5580.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.238233 -2.006360
## sample estimates:
## mean of x mean of y
## 22.80763 24.92992
```

Conclusion : The extremely small p-value ($p < 2.2e-16$) is much lower than our alpha level of significance ($\alpha = 0.05$) indicating that the difference in average ages between Master's and Ph.D. students is statistically significant. Therefore, we reject the null hypothesis, H_0 , which stated that we don't have enough evidence to conclude that the average of Master's students is equal to the average of PhD students. (2)

3. Compare the acceptance rate for students who learned about the university through family/friends versus those who found it through an internet search.

H_0 : Applicants who learnt about U of R through Family/Friend or Internet Search is independent of Acceptance status

H_A : Applicants who learnt about U of R through Family/Friend or Internet Search and Acceptance status are associated

```
##               Accepted NotAccepted
## FamilyFriends      851      1873
## InternetSearch     672      1677

##
## Pearson's Chi-squared test
##
## data:  chiMaster
## X-squared = 4.1619, df = 1, p-value = 0.04134

##               Accepted NotAccepted
## FamilyFriends       53       742
## InternetSearch      52       715

##
## Pearson's Chi-squared test
##
## data:  chiPhd
## X-squared = 0.0079489, df = 1, p-value = 0.929
```

$$\alpha = 0.05$$

Master Conclusion : Since the p-value is less than the significance level, we reject the null hypothesis and suggest that there is an association between how students learned about the university and the acceptance rate for the Master's Program.

PhD Conclusion : Since the p-value is greater than the significance level, we accept the null hypothesis and conclude that there is no association between how students learned about the university and the acceptance rate for the PhD program.

(3)

Logistic Regression:

1. Can we use factors such as age, citizenship, gender, recommender relationships, current employment status, and program choice to predict the likelihood of a candidate being accepted?

```
## [1] "program"          "degree"           "entry_term"
## [4] "time_status"      "sex"              "citizenship_status"
## [7] "failed_course"    "academic_probation" "inst_1_degree"
## [10] "heard_about_UR"

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##
## Call:
## glm(formula = decision_status ~ ., family = "binomial", data = train)
##
## Coefficients:
##                                     Estimate
## (Intercept)                        2.180e+00
## programData Science                 1.774e-01
## programElectrical and Computer Engineering 1.222e+00
## degreePhD                          -1.625e+00
## entry_termFall 2016                 1.603e-01
```

## entry_termFall 2017	-6.150e-03
## entry_termFall 2018	2.269e-01
## entry_termFall 2019	3.072e-02
## entry_termFall 2020	6.435e-02
## entry_termFall 2021	2.056e-01
## entry_termFall 2022	3.967e-01
## time_statusPart Time	7.884e-01
## time_statusunknown	-1.565e+01
## sexM	-2.288e-01
## sexunknown	-1.449e+01
## citizenship_statusPR	7.621e-01
## citizenship_statusunknown	-1.107e+01
## citizenship_statusUS	1.254e+00
## inst_1_degreeAssociate	-1.091e+00
## inst_1_degreeBA	-1.007e+00
## inst_1_degreeBachelor's	-1.048e+00
## inst_1_degreeBE	-8.895e-01
## inst_1_degreeBS	-5.696e-01
## inst_1_degreeCertification	-1.100e+00
## inst_1_degreeJD	-1.851e+01
## inst_1_degreeMA	-4.557e-01
## inst_1_degreeMaster's	1.296e-01
## inst_1_degreeMD	2.429e+00
## inst_1_degreeME	-1.348e+00
## inst_1_degreeMPH	-1.706e+01
## inst_1_degreeMS	-2.684e-01
## inst_1_degreeOther	-1.607e+00
## inst_1_degreeOther Bachelor's Degree	-9.831e-01
## inst_1_degreeOther Master's Degree	-7.746e-01
## inst_1_degreePhD	5.895e-01
## inst_1_degreeunknown	-9.801e-01
## heard_about_URAttended a talk by a faculty member from the University of Rochester	9.044e-01
## heard_about_URFamily/Friend	2.444e-02
## heard_about_URInternet search	-1.678e-01
## heard_about_URLocal resident (current or past)	3.054e-01
## heard_about_URMailing	-3.736e-01
## heard_about_UROther	1.079e-01
## heard_about_URPrevious participant in a program at the University of Rochester	7.747e-01
## heard_about_URRead a publication by a faculty member from the University of Rochester	3.266e-01
## heard_about_URRecruitment email	-3.460e-01
## heard_about_URRecruitment/Graduate School fair	1.741e-01
## heard_about_URReferral from an affiliate of the University of Rochester	2.368e-01
## heard_about_URReferral from someone at my school (mentor/advisor/faculty)	1.705e-01
## heard_about_URThird-party website (e.g. gradschools.com, Peterson's, etc.)	3.843e-02
## heard_about_URUniversity of Rochester reputation/word of mouth	1.830e-01
## heard_about_URunknown	-1.291e+00
## age	-1.234e-01
## failed_course1	-4.356e-01
## failed_courseunknown	-1.405e+01
## academic_probation1	-3.583e-02
## academic_probationunknown	-1.309e+01
##	Std. Error
## (Intercept)	1.082e+00
## programData Science	6.775e-02

## programElectrical and Computer Engineering	6.095e-02
## degreePhD	9.131e-02
## entry_termFall 2016	2.180e-01
## entry_termFall 2017	2.200e-01
## entry_termFall 2018	2.180e-01
## entry_termFall 2019	2.183e-01
## entry_termFall 2020	2.177e-01
## entry_termFall 2021	2.199e-01
## entry_termFall 2022	2.180e-01
## time_statusPart Time	2.807e-01
## time_statusunknown	1.886e+03
## sexM	5.625e-02
## sexunknown	3.754e+02
## citizenship_statusPR	2.186e-01
## citizenship_statusunknown	3.522e+02
## citizenship_statusUS	9.624e-02
## inst_1_degreeAssociate	1.326e+00
## inst_1_degreeBA	8.144e-01
## inst_1_degreeBachelor's	8.366e-01
## inst_1_degreeBE	8.062e-01
## inst_1_degreeBS	8.053e-01
## inst_1_degreeCertification	9.991e-01
## inst_1_degreeJD	1.075e+04
## inst_1_degreeMA	9.329e-01
## inst_1_degreeMaster's	8.608e-01
## inst_1_degreeMD	1.644e+00
## inst_1_degreeME	8.663e-01
## inst_1_degreeMPH	5.331e+03
## inst_1_degreeMS	8.097e-01
## inst_1_degreeOther	1.014e+00
## inst_1_degreeOther Bachelor's Degree	8.086e-01
## inst_1_degreeOther Master's Degree	8.691e-01
## inst_1_degreePhD	8.813e-01
## inst_1_degreeunknown	8.076e-01
## heard_about_URAttended a talk by a faculty member from the University of Rochester	6.919e-01
## heard_about_URFamily/Friend	6.295e-01
## heard_about_URInternet search	6.301e-01
## heard_about_URLocal resident (current or past)	6.850e-01
## heard_about_URMailing	6.628e-01
## heard_about_UROther	6.450e-01
## heard_about_URPrevious participant in a program at the University of Rochester	6.503e-01
## heard_about_URRead a publication by a faculty member from the University of Rochester	6.533e-01
## heard_about_URRecruitment email	6.587e-01
## heard_about_URRecruitment/Graduate School fair	7.715e-01
## heard_about_URReferral from an affiliate of the University of Rochester	6.999e-01
## heard_about_URReferral from someone at my school (mentor/advisor/faculty)	6.343e-01
## heard_about_URThird-party website (e.g. gradschools.com, Peterson's, etc.)	6.511e-01
## heard_about_URUniversity of Rochester reputation/word of mouth	6.296e-01
## heard_about_URunknown	6.429e-01
## age	1.249e-02
## failed_course1	7.751e-02
## failed_courseunknown	3.193e+02
## academic_probation1	2.649e-01
## academic_probationunknown	3.202e+02

##	z value
## (Intercept)	2.014
## programData Science	2.619
## programElectrical and Computer Engineering	20.051
## degreePhD	-17.800
## entry_termFall 2016	0.735
## entry_termFall 2017	-0.028
## entry_termFall 2018	1.041
## entry_termFall 2019	0.141
## entry_termFall 2020	0.296
## entry_termFall 2021	0.935
## entry_termFall 2022	1.819
## time_statusPart Time	2.809
## time_statusunknown	-0.008
## sexM	-4.067
## sexunknown	-0.039
## citizenship_statusPR	3.486
## citizenship_statusunknown	-0.031
## citizenship_statusUS	13.029
## inst_1_degreeAssociate	-0.823
## inst_1_degreeBA	-1.237
## inst_1_degreeBachelor's	-1.253
## inst_1_degreeBE	-1.103
## inst_1_degreeBS	-0.707
## inst_1_degreeCertification	-1.101
## inst_1_degreeJD	-0.002
## inst_1_degreeMA	-0.488
## inst_1_degreeMaster's	0.151
## inst_1_degreeMD	1.478
## inst_1_degreeME	-1.555
## inst_1_degreeMPH	-0.003
## inst_1_degreeMS	-0.332
## inst_1_degreeOther	-1.584
## inst_1_degreeOther Bachelor's Degree	-1.216
## inst_1_degreeOther Master's Degree	-0.891
## inst_1_degreePhD	0.669
## inst_1_degreeunknown	-1.214
## heard_about_URAttended a talk by a faculty member from the University of Rochester	1.307
## heard_about_URFamily/Friend	0.039
## heard_about_URInternet search	-0.266
## heard_about_URLocal resident (current or past)	0.446
## heard_about_URMailing	-0.564
## heard_about_UROther	0.167
## heard_about_URPrevious participant in a program at the University of Rochester	1.191
## heard_about_URRead a publication by a faculty member from the University of Rochester	0.500
## heard_about_URRecruitment email	-0.525
## heard_about_URRecruitment/Graduate School fair	0.226
## heard_about_URReferral from an affiliate of the University of Rochester	0.338
## heard_about_URReferral from someone at my school (mentor/advisor/faculty)	0.269
## heard_about_URThird-party website (e.g. gradschools.com, Peterson's, etc.)	0.059
## heard_about_URUniversity of Rochester reputation/word of mouth	0.291
## heard_about_URunknown	-2.008
## age	-9.878
## failed_course1	-5.620

## failed_courseunknown	-0.044
## academic_probation1	-0.135
## academic_probationunknown	-0.041
##	Pr(> z)
## (Intercept)	0.044058
## programData Science	0.008829
## programElectrical and Computer Engineering	< 2e-16
## degreePhD	< 2e-16
## entry_termFall 2016	0.462049
## entry_termFall 2017	0.977694
## entry_termFall 2018	0.298000
## entry_termFall 2019	0.888080
## entry_termFall 2020	0.767586
## entry_termFall 2021	0.349732
## entry_termFall 2022	0.068872
## time_statusPart Time	0.004975
## time_statusunknown	0.993379
## sexM	4.76e-05
## sexunknown	0.969201
## citizenship_statusPR	0.000489
## citizenship_statusunknown	0.974935
## citizenship_statusUS	< 2e-16
## inst_1_degreeAssociate	0.410683
## inst_1_degreeBA	0.216268
## inst_1_degreeBachelor's	0.210275
## inst_1_degreeBE	0.269890
## inst_1_degreeBS	0.479406
## inst_1_degreeCertification	0.271010
## inst_1_degreeJD	0.998627
## inst_1_degreeMA	0.625219
## inst_1_degreeMaster's	0.880347
## inst_1_degreeMD	0.139530
## inst_1_degreeME	0.119834
## inst_1_degreeMPH	0.997447
## inst_1_degreeMS	0.740239
## inst_1_degreeOther	0.113149
## inst_1_degreeOther Bachelor's Degree	0.224070
## inst_1_degreeOther Master's Degree	0.372786
## inst_1_degreePhD	0.503554
## inst_1_degreeunknown	0.224890
## heard_about_URAttended a talk by a faculty member from the University of Rochester	0.191157
## heard_about_URFamily/Friend	0.969028
## heard_about_URInternet search	0.790044
## heard_about_URLocal resident (current or past)	0.655733
## heard_about_URMailing	0.573030
## heard_about_UROther	0.867186
## heard_about_URPrevious participant in a program at the University of Rochester	0.233553
## heard_about_URRead a publication by a faculty member from the University of Rochester	0.617155
## heard_about_URRecruitment email	0.599374
## heard_about_URRecruitment/Graduate School fair	0.821515
## heard_about_URReferral from an affiliate of the University of Rochester	0.735146
## heard_about_URReferral from someone at my school (mentor/advisor/faculty)	0.788118
## heard_about_URThird-party website (e.g. gradschools.com, Peterson's, etc.)	0.952927
## heard_about_URUniversity of Rochester reputation/word of mouth	0.771284

```

## heard_about_URunknown 0.044665
## age < 2e-16
## failed_course1 1.91e-08
## failed_courseunknown 0.964892
## academic_probation1 0.892409
## academic_probationunknown 0.967376
##
## (Intercept) *
## programData Science **
## programElectrical and Computer Engineering ***
## degreePhD ***
## entry_termFall 2016
## entry_termFall 2017
## entry_termFall 2018
## entry_termFall 2019
## entry_termFall 2020
## entry_termFall 2021
## entry_termFall 2022 .
## time_statusPart Time **
## time_statusunknown
## sexM ***
## sexunknown
## citizenship_statusPR ***
## citizenship_statusunknown
## citizenship_statusUS ***
## inst_1_degreeAssociate
## inst_1_degreeBA
## inst_1_degreeBachelor's
## inst_1_degreeBE
## inst_1_degreeBS
## inst_1_degreeCertification
## inst_1_degreeJD
## inst_1_degreeMA
## inst_1_degreeMaster's
## inst_1_degreeMD
## inst_1_degreeME
## inst_1_degreeMPH
## inst_1_degreeMS
## inst_1_degreeOther
## inst_1_degreeOther Bachelor's Degree
## inst_1_degreeOther Master's Degree
## inst_1_degreePhD
## inst_1_degreeunknown
## heard_about_URAttended a talk by a faculty member from the University of Rochester
## heard_about_URFamily/Friend
## heard_about_URInternet search
## heard_about_URLocal resident (current or past)
## heard_about_URMailing
## heard_about_UROther
## heard_about_URPrevious participant in a program at the University of Rochester
## heard_about_URRead a publication by a faculty member from the University of Rochester
## heard_about_URRecruitment email
## heard_about_URRecruitment/Graduate School fair
## heard_about_URReferral from an affiliate of the University of Rochester

```

```

## heard_about_URReferral from someone at my school (mentor/advisor/faculty)
## heard_about_URThird-party website (e.g.  gradschoools.com, Peterson's, etc.)
## heard_about_URUniversity of Rochester reputation/word of mouth
## heard_about_URunknown
## age
## failed_course1
## failed_courseunknown
## academic_probation1
## academic_probationunknown
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 11763.3  on 11350  degrees of freedom
## Residual deviance:  9786.6  on 11295  degrees of freedom
## AIC: 9898.6
##
## Number of Fisher Scoring iterations: 18

##      Actual
## Predicted    0    1
##      0 3673  824
##      1  154  213

## [1] "Accuracy: 0.798930921052632"

```