

Screen background data

Dominik Glandorf

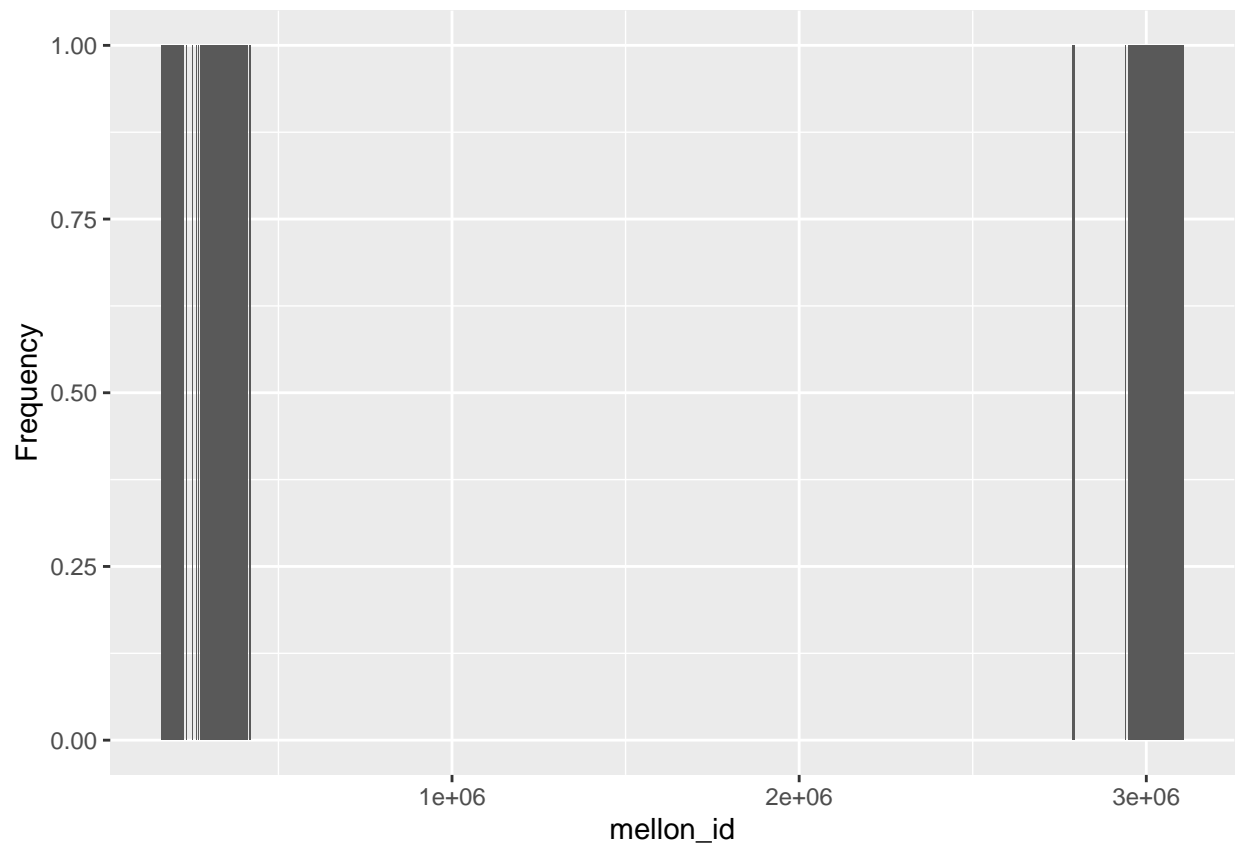
2022-11-22

Background variables

The spreadsheet knows 161 variables. The corresponding file contains 199 columns. The expected variables freshman,transfer are not in the dataset. Therefore exist variables pascd,hscd,must_hsid,ncessch,hs_address,hs_city,hs_state

Available variables

```
[1] Variable: mellon_id, type: numeric, requested: yes
[1] Values (97049 unique): 399229, 196276, 185813, 173223, 190267, ...
[1] Missing: 0%
```



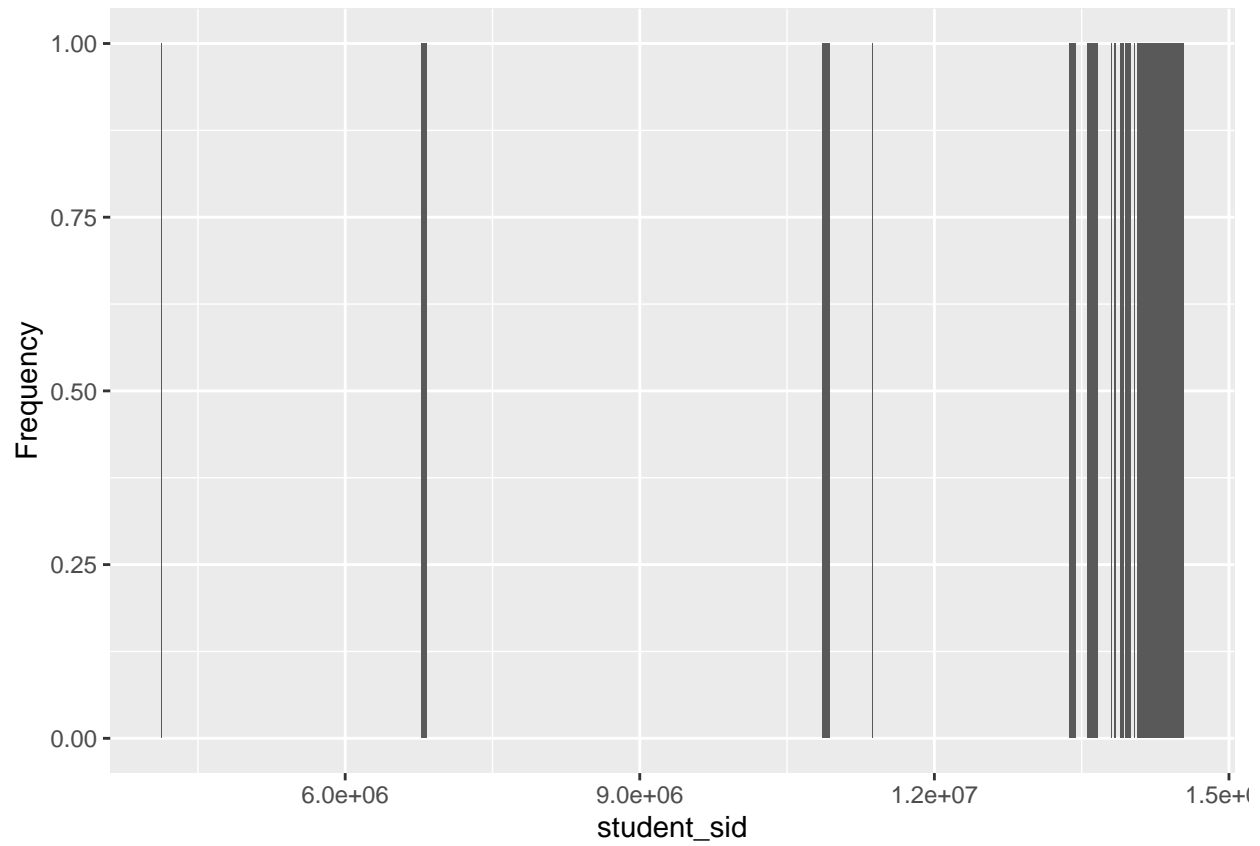
```
[1] should not be used as a predictor
```

```
[1] -----
```

```
[1] Variable: student_sid, type: numeric, requested: yes
```

```
[1] Values (97049 unique): 4128117, 10857014, 10857958, 10859282, 10863615, ...
```

```
[1] Missing: 0%
```



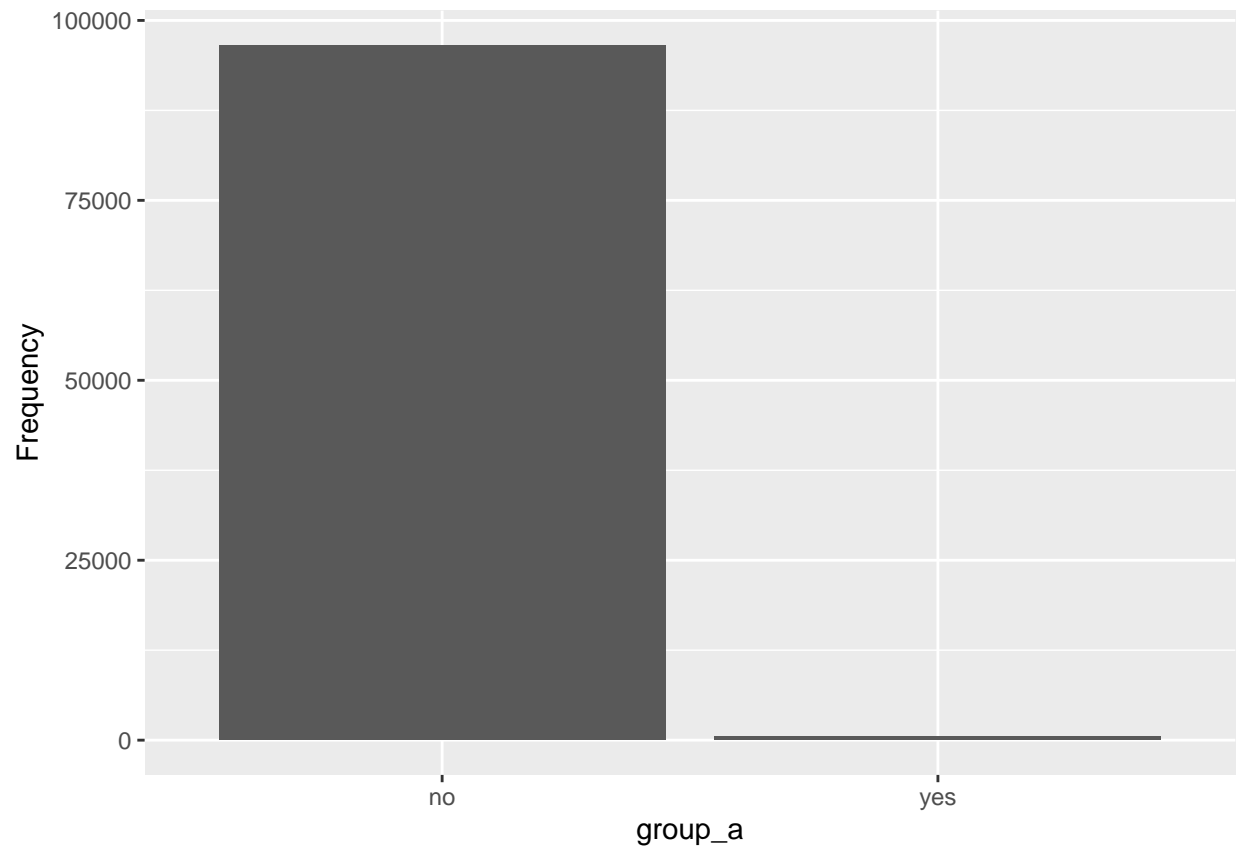
```
[1] should not be used as a predictor
```

```
[1] -----
```

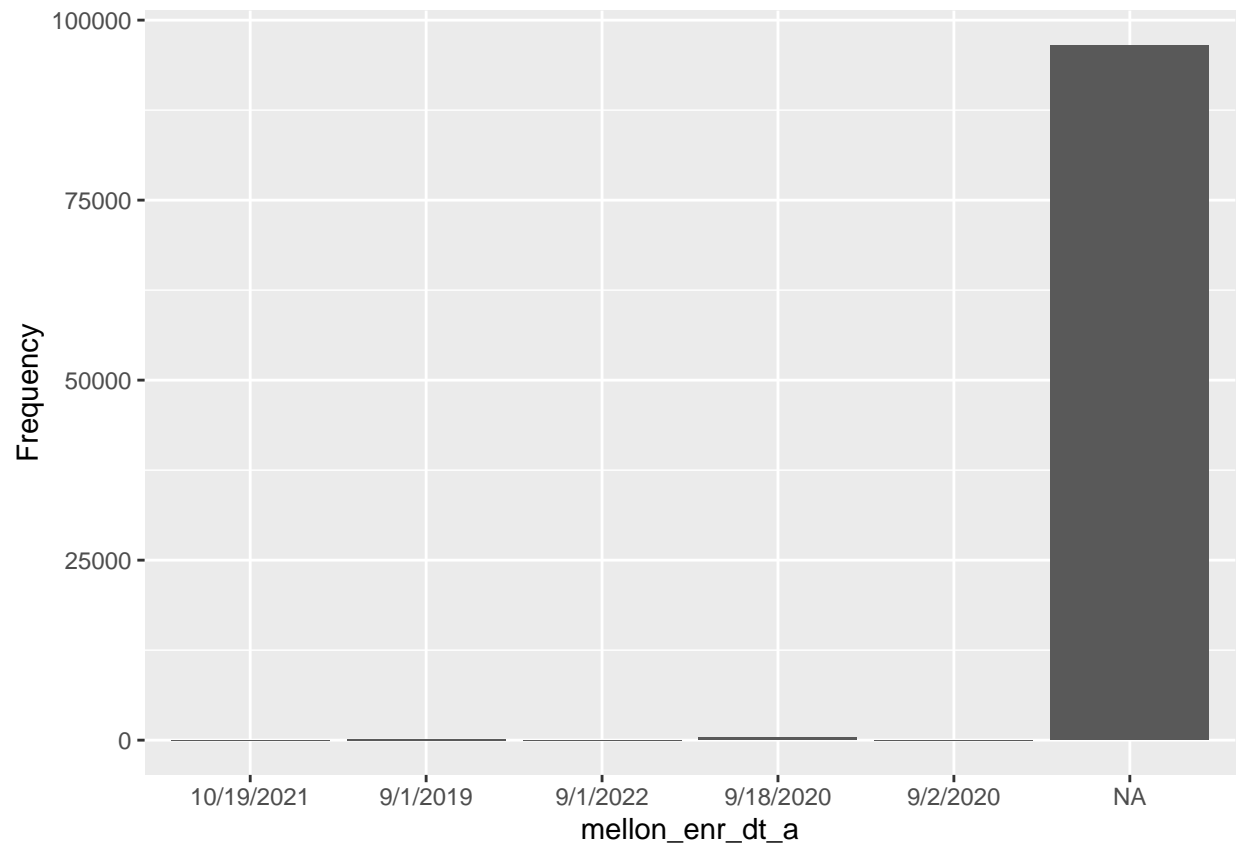
```
[1] Variable: group_a, type: character, requested: yes
```

```
[1] Values (2 unique): no, yes
```

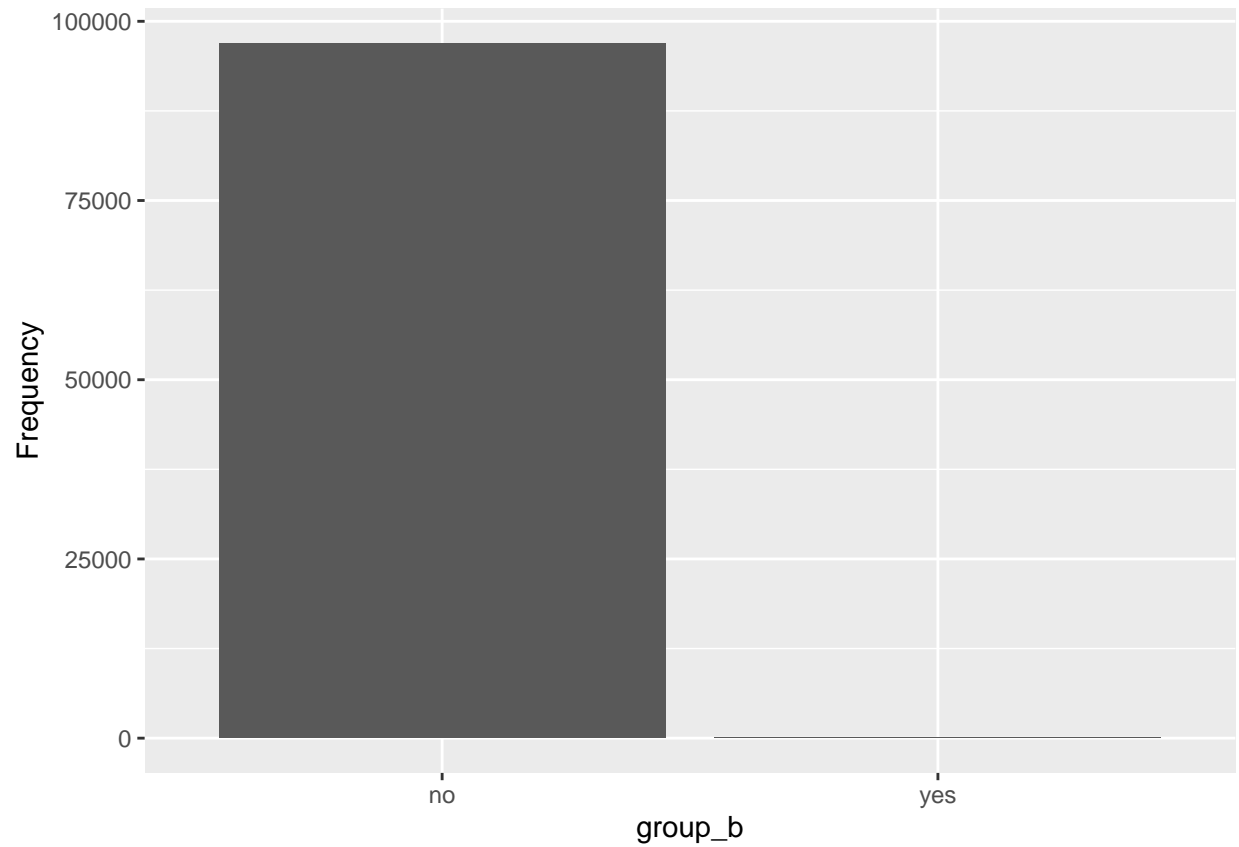
```
[1] Missing: 0%
```



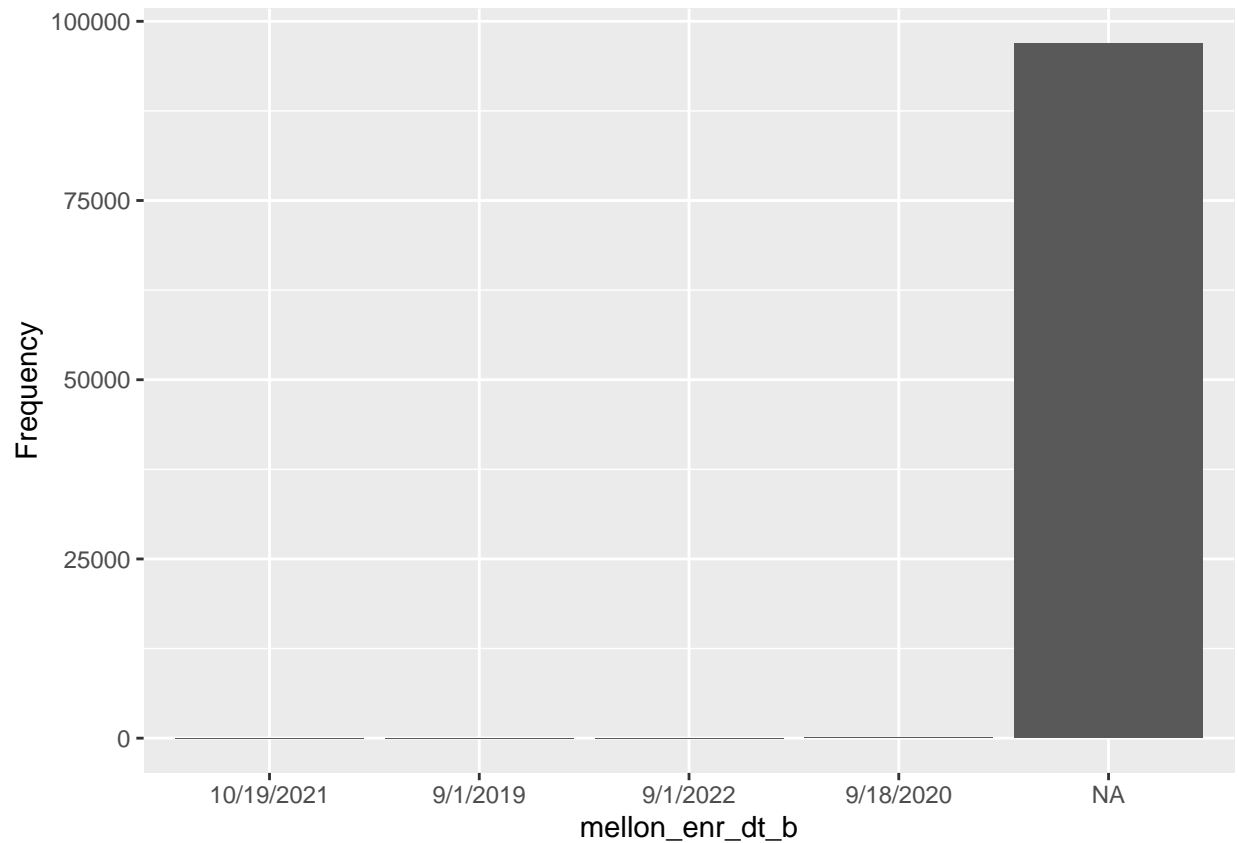
```
[1] should not be used as a predictor
[1] -----
[1] Variable: mellon_enr_dt_a, type: character, requested: yes
[1] Values (6 unique): NA, 9/18/2020, 9/1/2019, 10/19/2021, 9/1/2022, ...
[1] Missing: 99.4%
```



```
[1] should not be used as a predictor
[1] -----
[1] Variable: group_b, type: character, requested: yes
[1] Values (2 unique): no, yes
[1] Missing: 0%
```

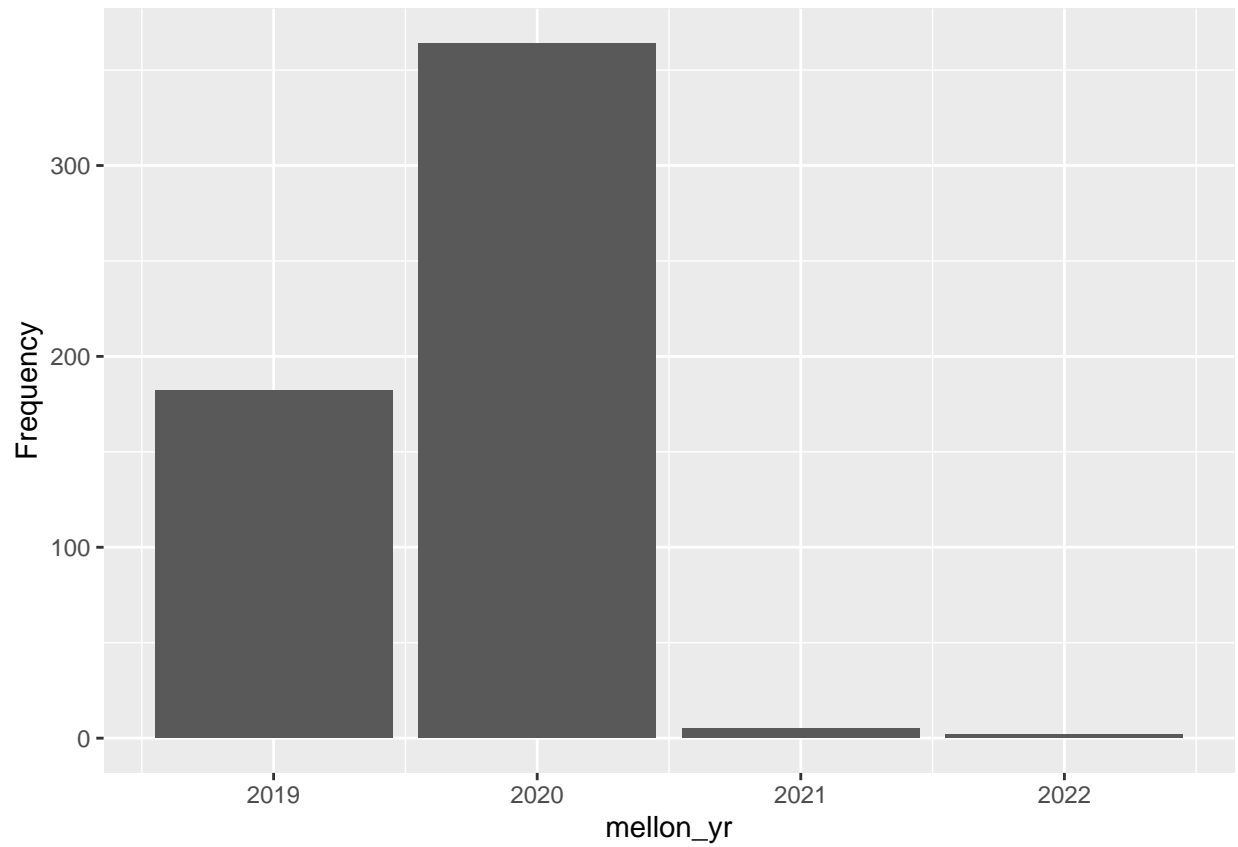


```
[1] should not be used as a predictor
[1] -----
[1] Variable: mellon_enr_dt_b, type: character, requested: yes
[1] Values (5 unique): NA, 9/18/2020, 9/1/2019, 10/19/2021, 9/1/2022
[1] Missing: 99.9%
```

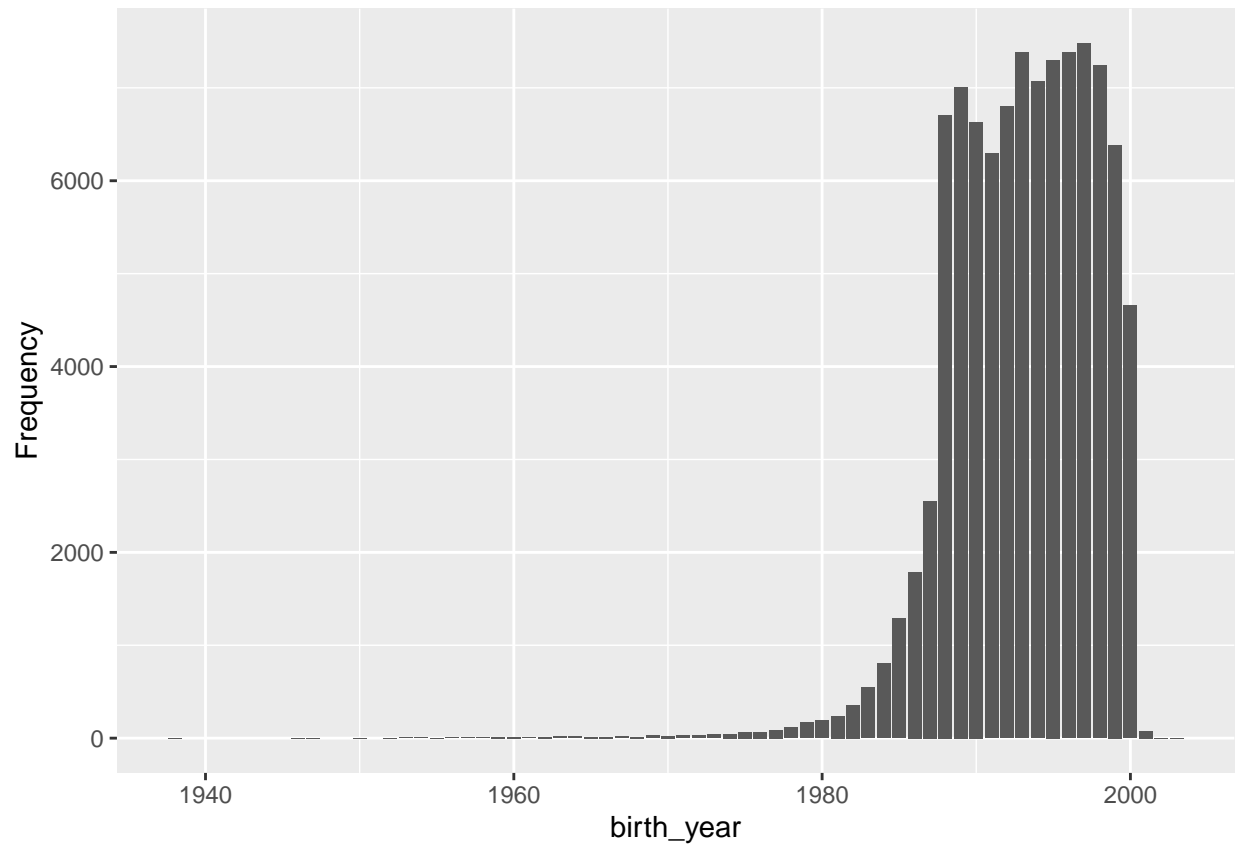


```
[1] should not be used as a predictor
[1] -----
[1] Variable: mellon_yr, type: numeric, requested: yes
[1] Values (5 unique): NA, 2020, 2019, 2021, 2022
[1] Missing: 99.4%
```

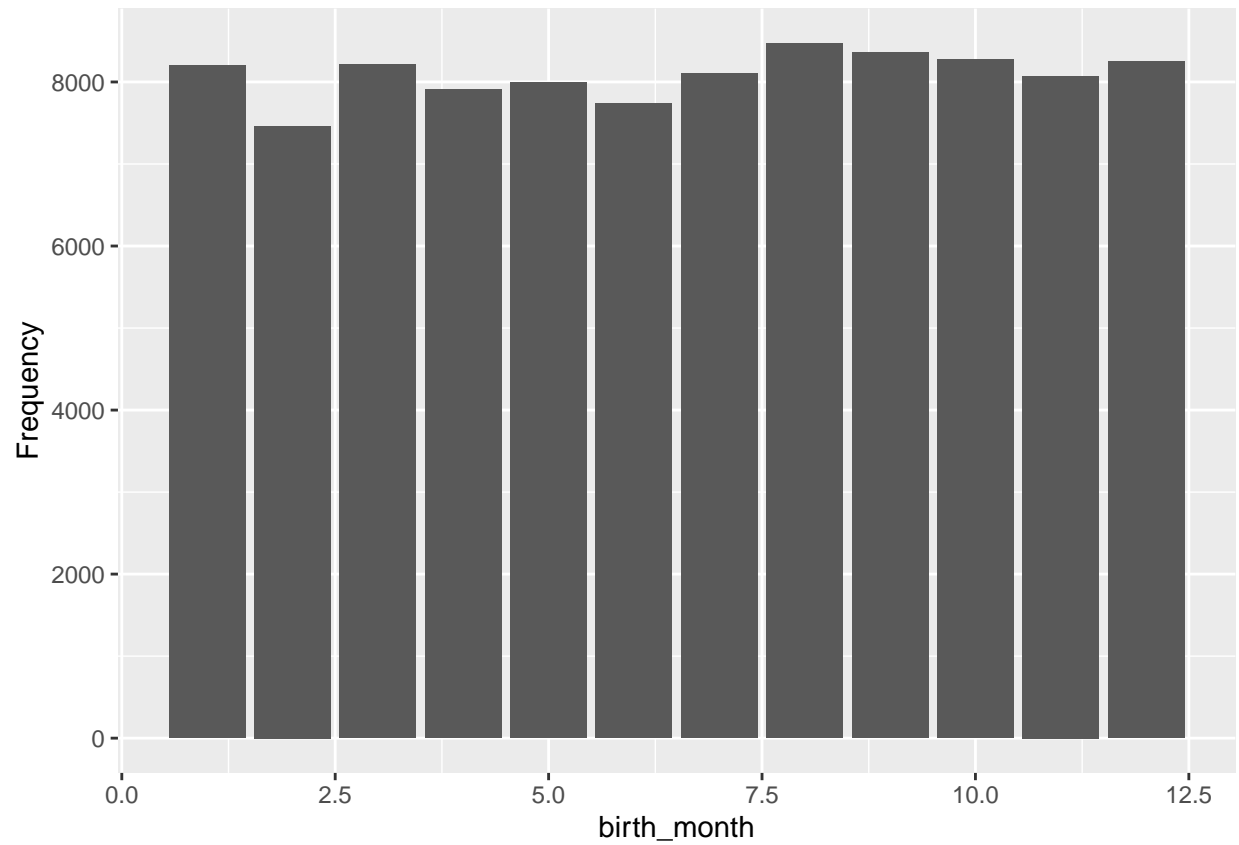
Warning: Removed 96496 rows containing non-finite values ('stat_count()').



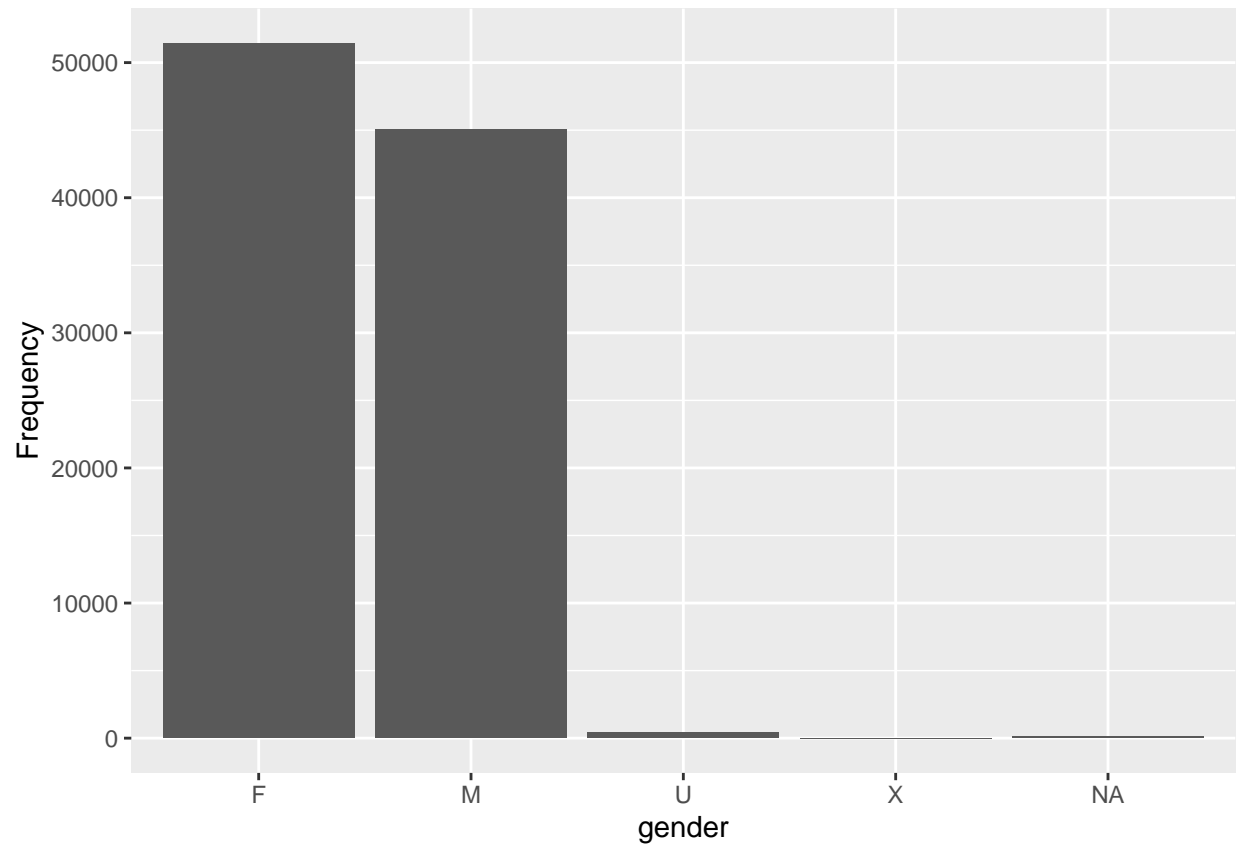
```
[1] should not be used as a predictor
[1] -----
[1] Variable: birth_year, type: numeric, requested: yes
[1] Values (56 unique): 1977, 1997, 2000, 1999, 1995, ...
[1] Missing: 0%
```



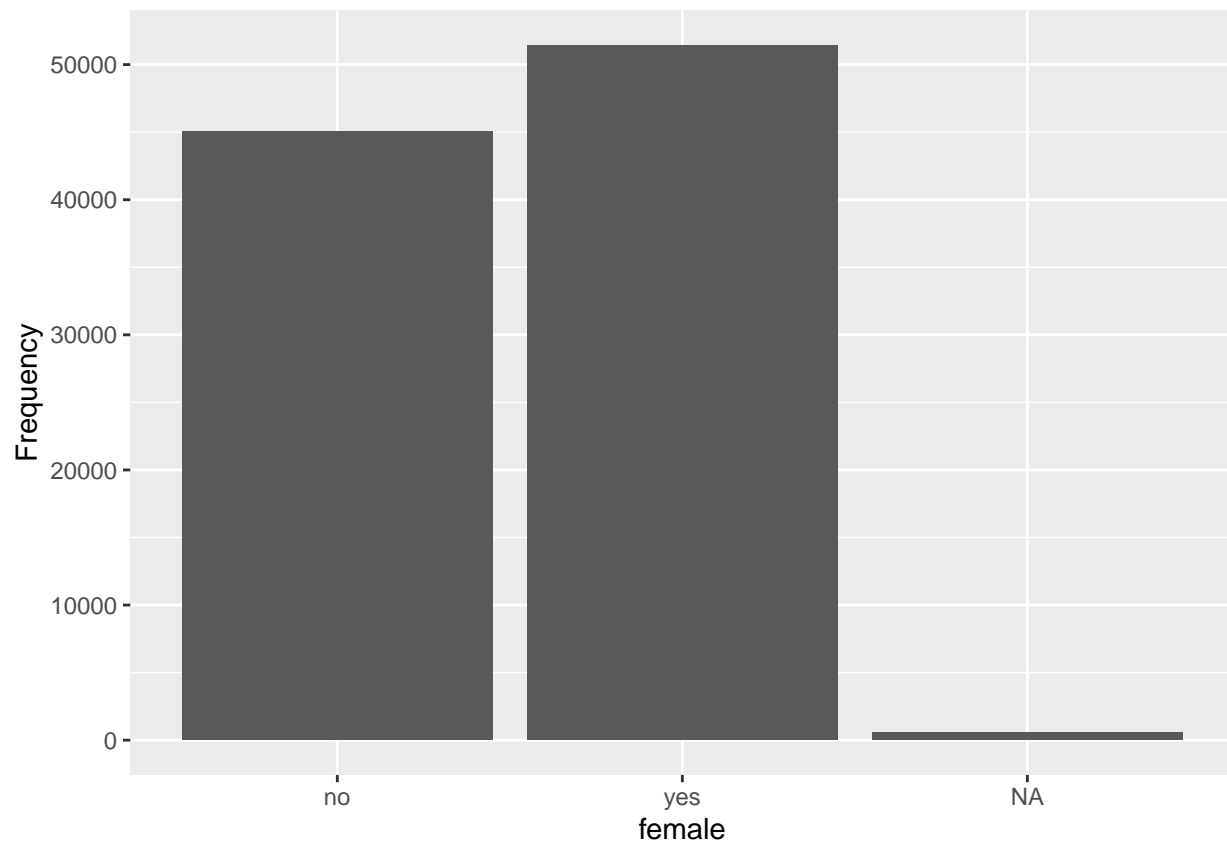
```
[1] is used in feature engineering
[1] -----
[1] Variable: birth_month, type: numeric, requested: yes
[1] Values (12 unique): 3, 11, 1, 7, 8, ...
[1] Missing: 0%
```

```
[1] is used in feature engineering
[1] -----
[1] Variable: gender, type: character, requested: yes
[1] Values (5 unique): M, F, U, NA, X
[1] Missing: 0.1%
```

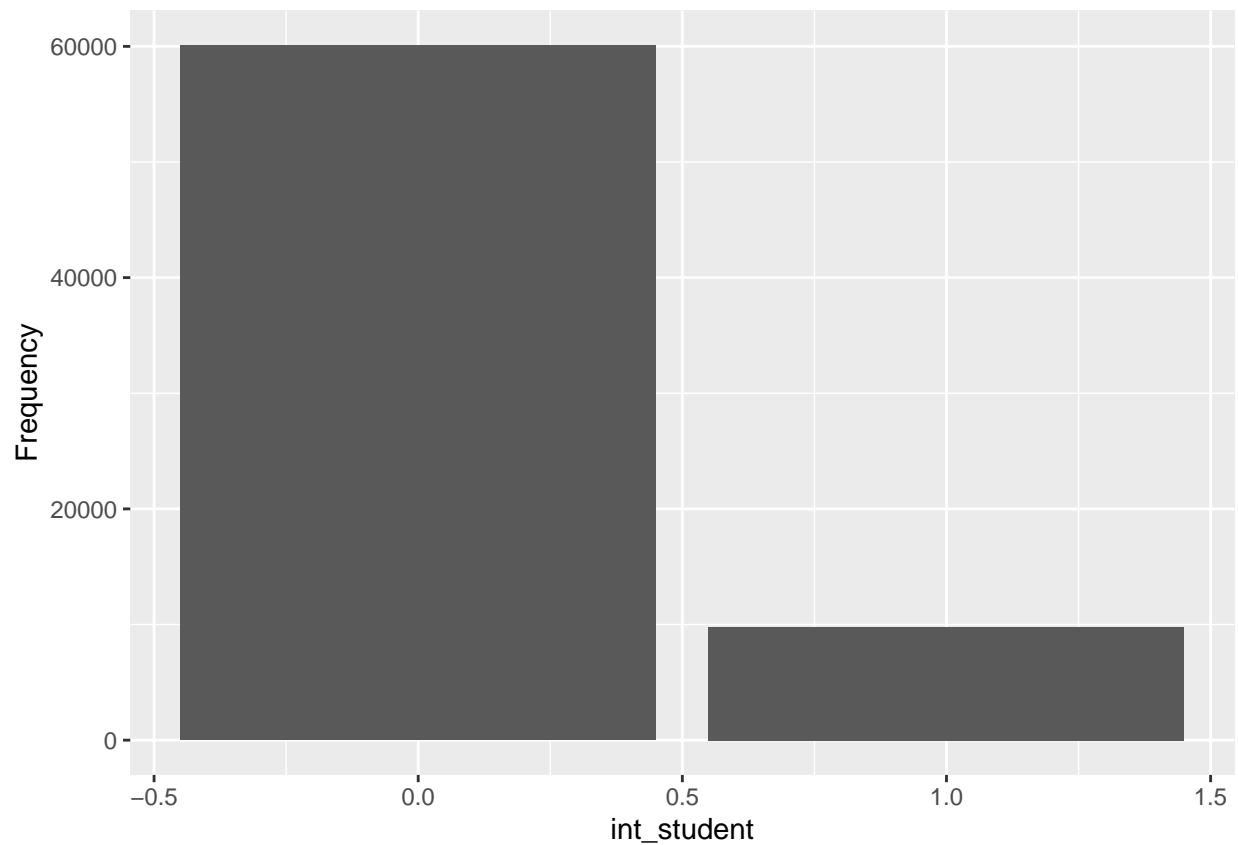


```
[1] -----  
[1] Variable: female, type: character, requested: yes  
[1] Values (3 unique): no, yes, NA  
[1] Missing: 0.6%
```

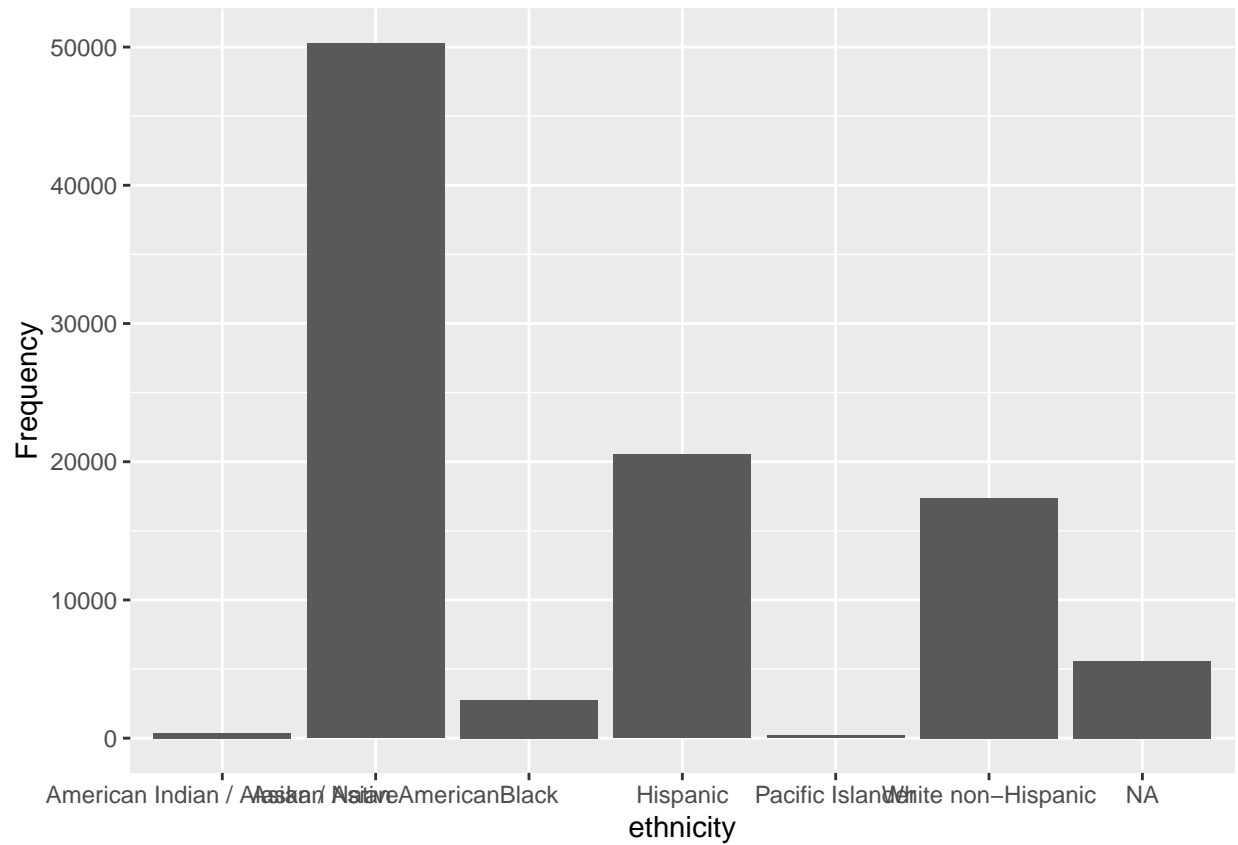


```
[1] -----
[1] Variable: int_student, type: numeric, requested: yes
[1] Values (3 unique): NA, 0, 1
[1] Missing: 28%
  Group.1 int_student
1    2006  0.99701961
2    2007  0.99749447
3    2008  0.99630826
4    2009  0.99496560
5    2010  0.02982537
6    2011  0.02932011
7    2012  0.02499282
8    2013  0.02439664
9    2014  0.03190104
10   2015  0.03164397
11   2016  0.02497578
12   2017  0.01512813
13   2018  0.00000000
```

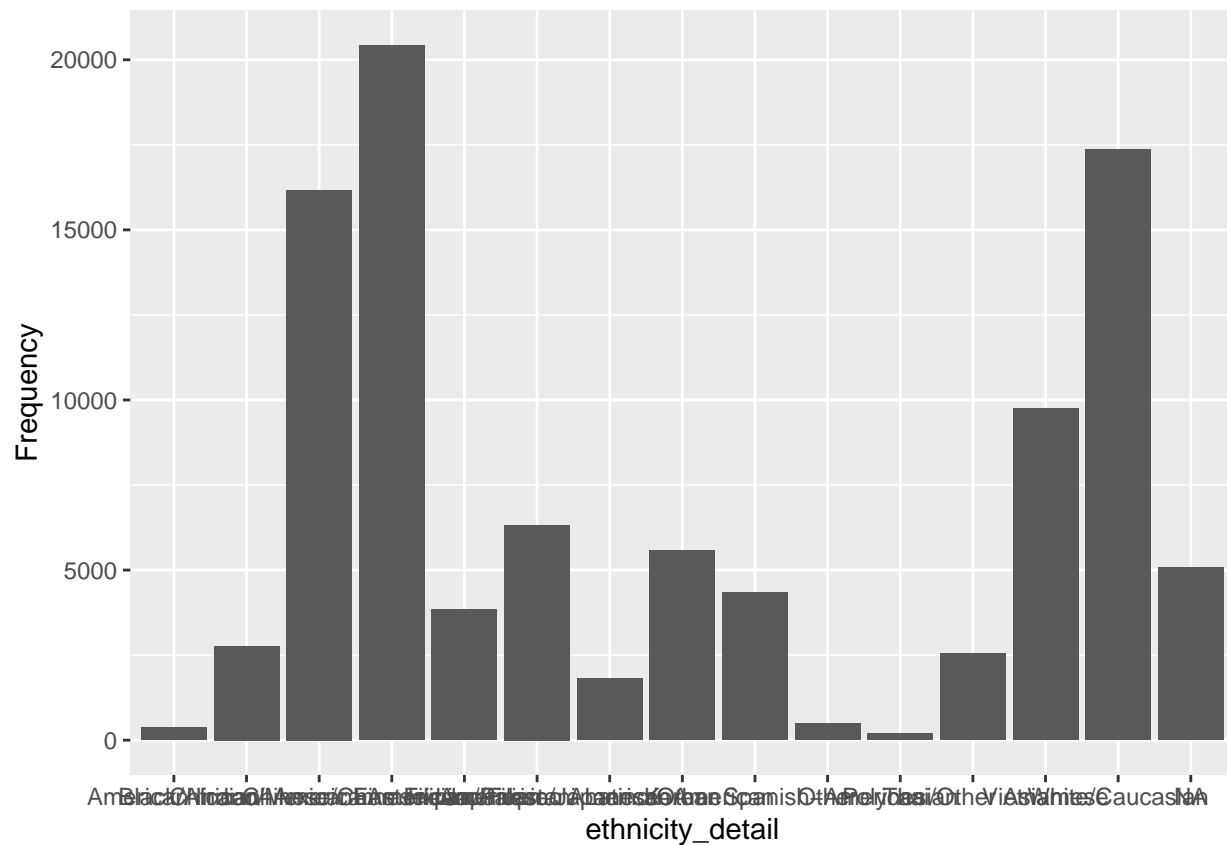
Warning: Removed 27165 rows containing non-finite values ('stat_count()').



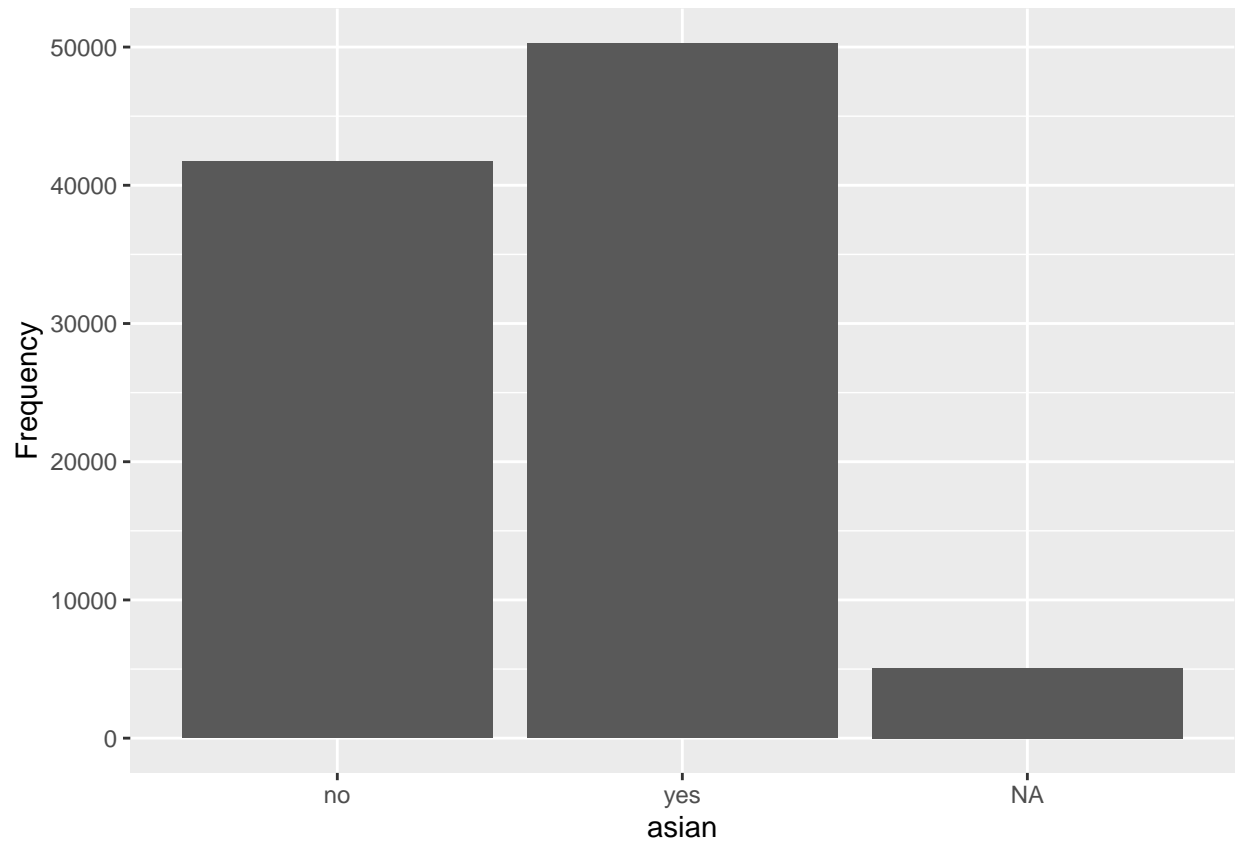
```
[1] -----  
[1] Variable: ethnicity, type: character, requested: yes  
[1] Values (7 unique): White non-Hispanic, NA, Asian / Asian American, Hispanic, Black, ...  
[1] Missing: 5.8%  
[1] Most missing: 2008 8.2%, Least missing: 2012 4.3%
```



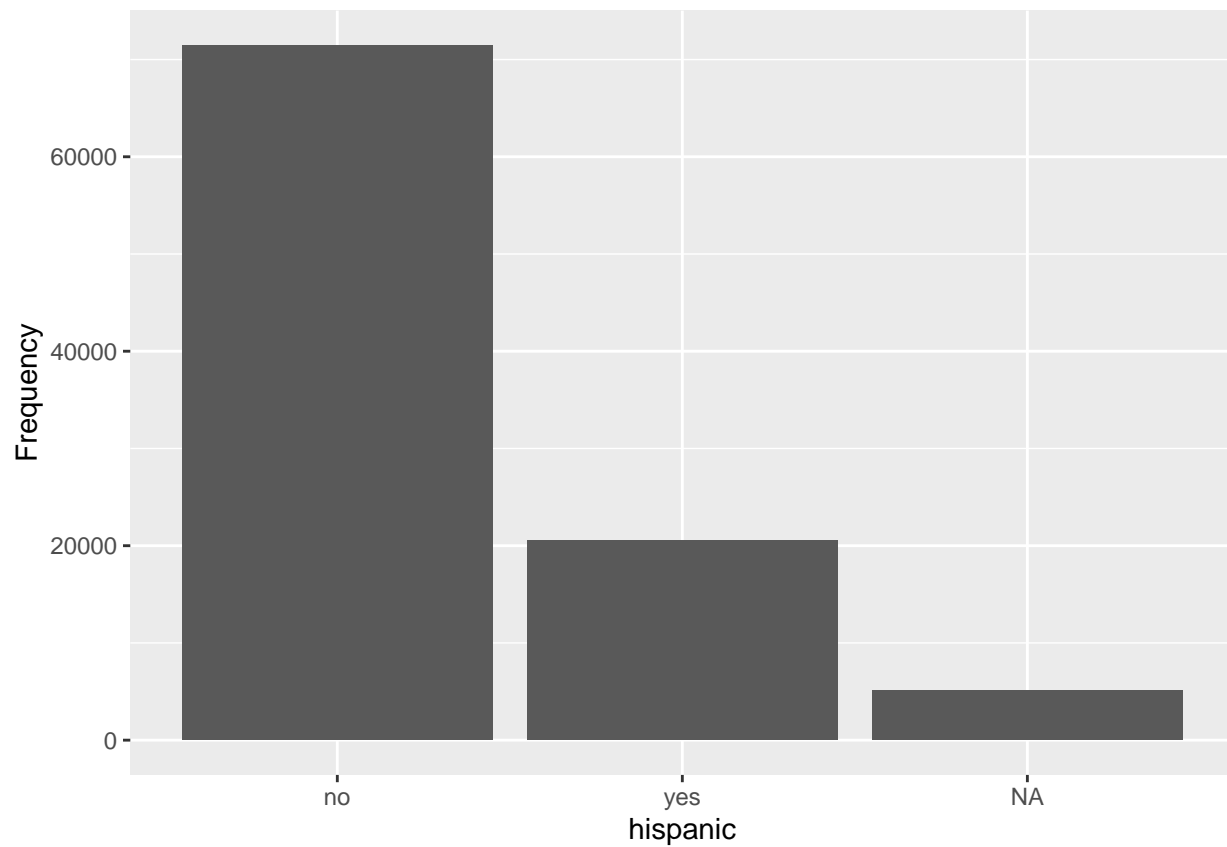
```
[1] -----
[1] Variable: ethnicity_detail, type: character, requested: yes
[1] Values (15 unique): White/Caucasian, NA, East Indian/Pakistani, Chicano/Mexican-American, Korean, .
[1] Missing: 5.2%
[1] Most missing: 2010 7.8%, Least missing: 2006 4%
```



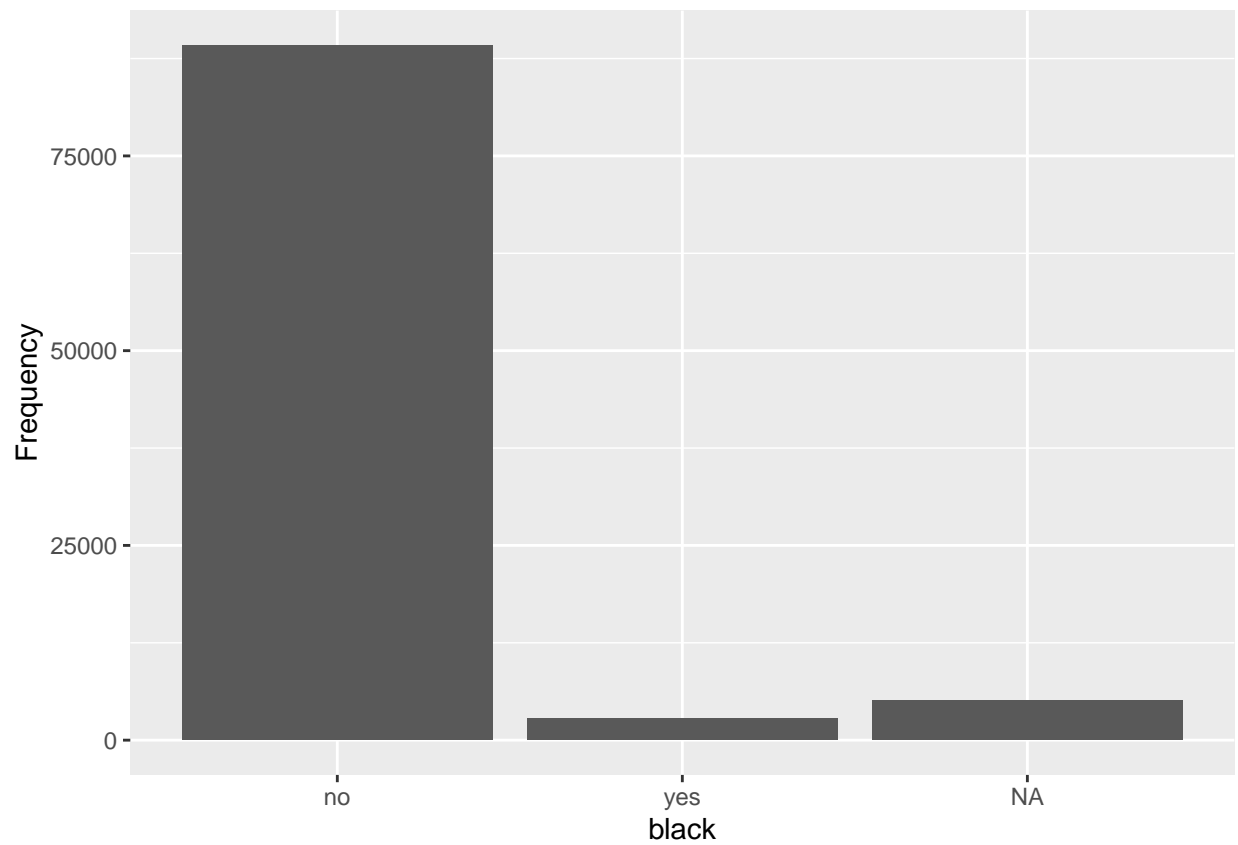
```
[1] -----
[1] Variable: asian, type: character, requested: yes
[1] Values (3 unique): no, NA, yes
[1] Missing: 5.2%
[1] Most missing: 2010 7.8%, Least missing: 2006 4%
```



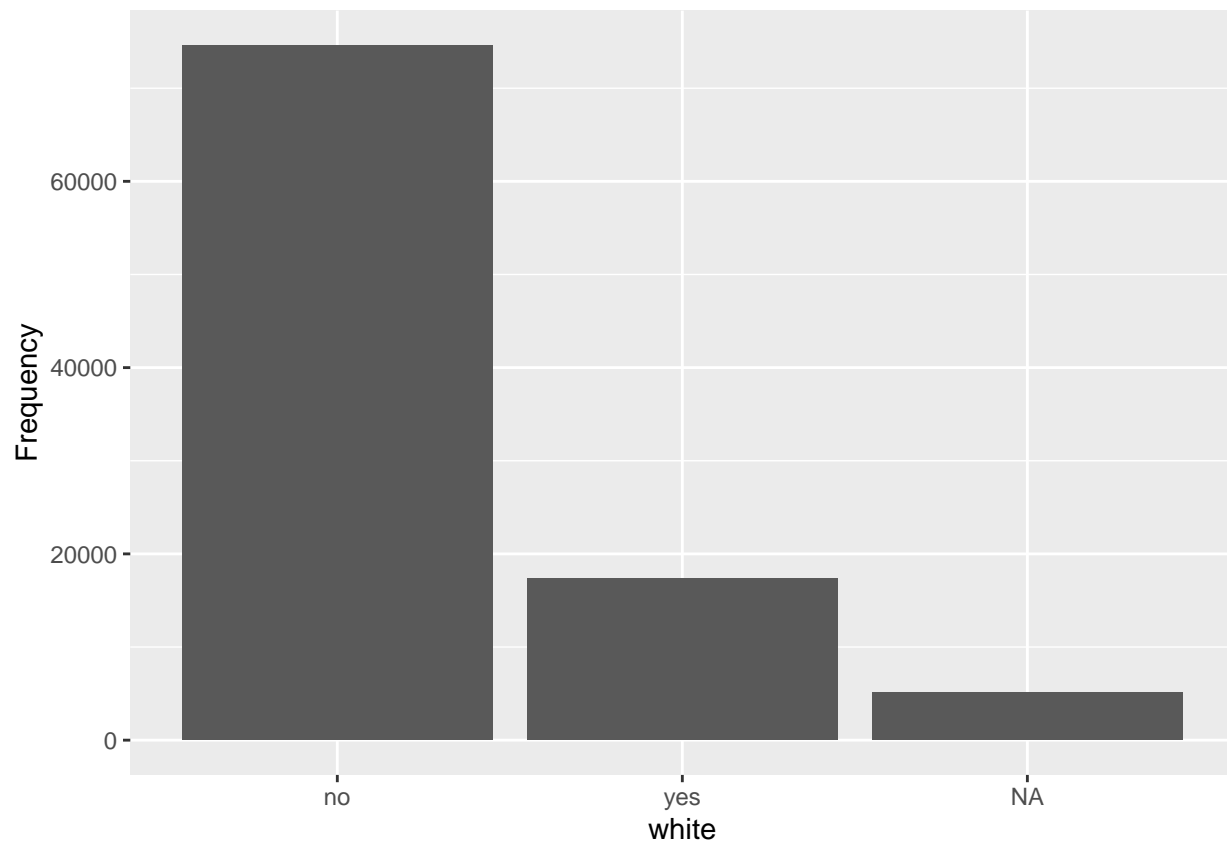
```
[1] -----  
[1] Variable: hispanic, type: character, requested: yes  
[1] Values (3 unique): no, NA, yes  
[1] Missing: 5.2%  
[1] Most missing: 2010 7.8%, Least missing: 2006 4%
```



```
[1] -----  
[1] Variable: black, type: character, requested: yes  
[1] Values (3 unique): no, NA, yes  
[1] Missing: 5.2%  
[1] Most missing: 2010 7.8%, Least missing: 2006 4%
```

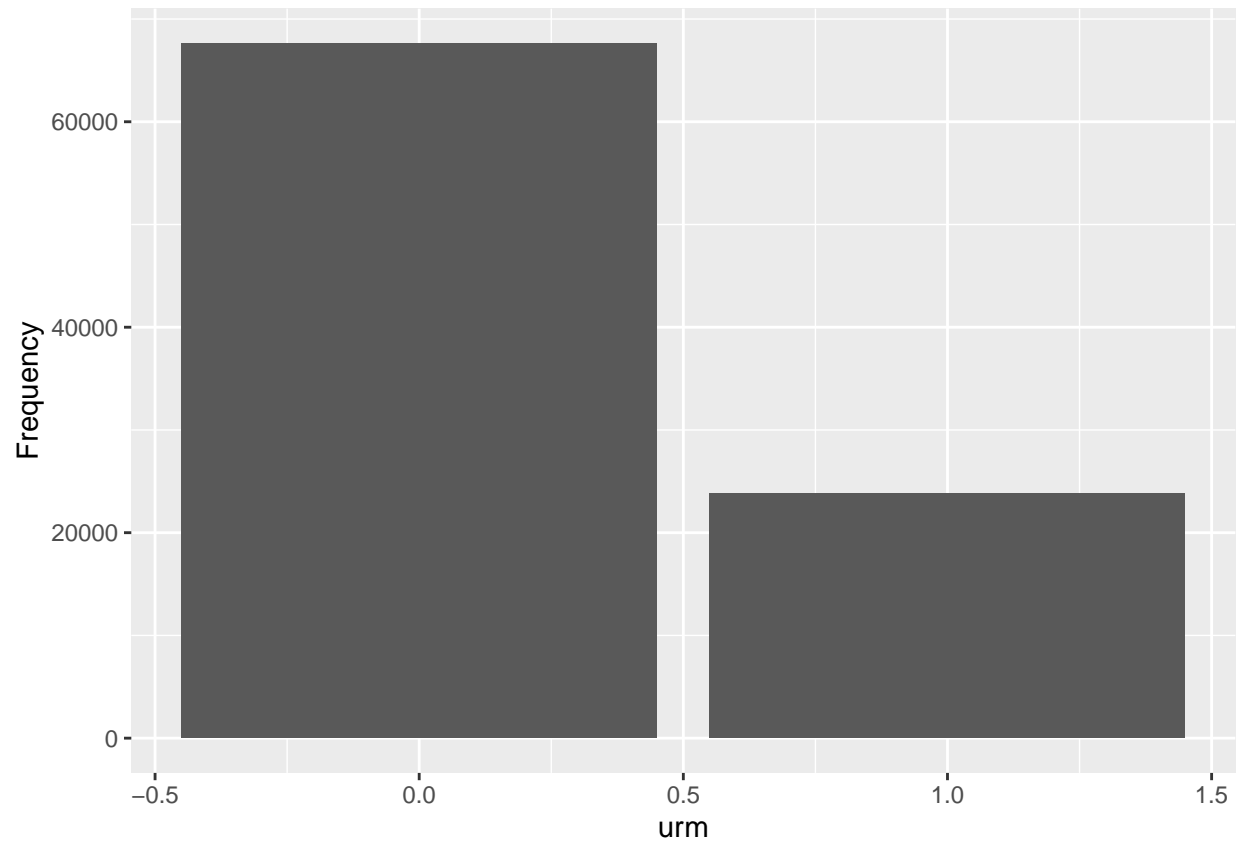



```
[1] -----  
[1] Variable: white, type: character, requested: yes  
[1] Values (3 unique): yes, NA, no  
[1] Missing: 5.2%  
[1] Most missing: 2010 7.8%, Least missing: 2006 4%
```

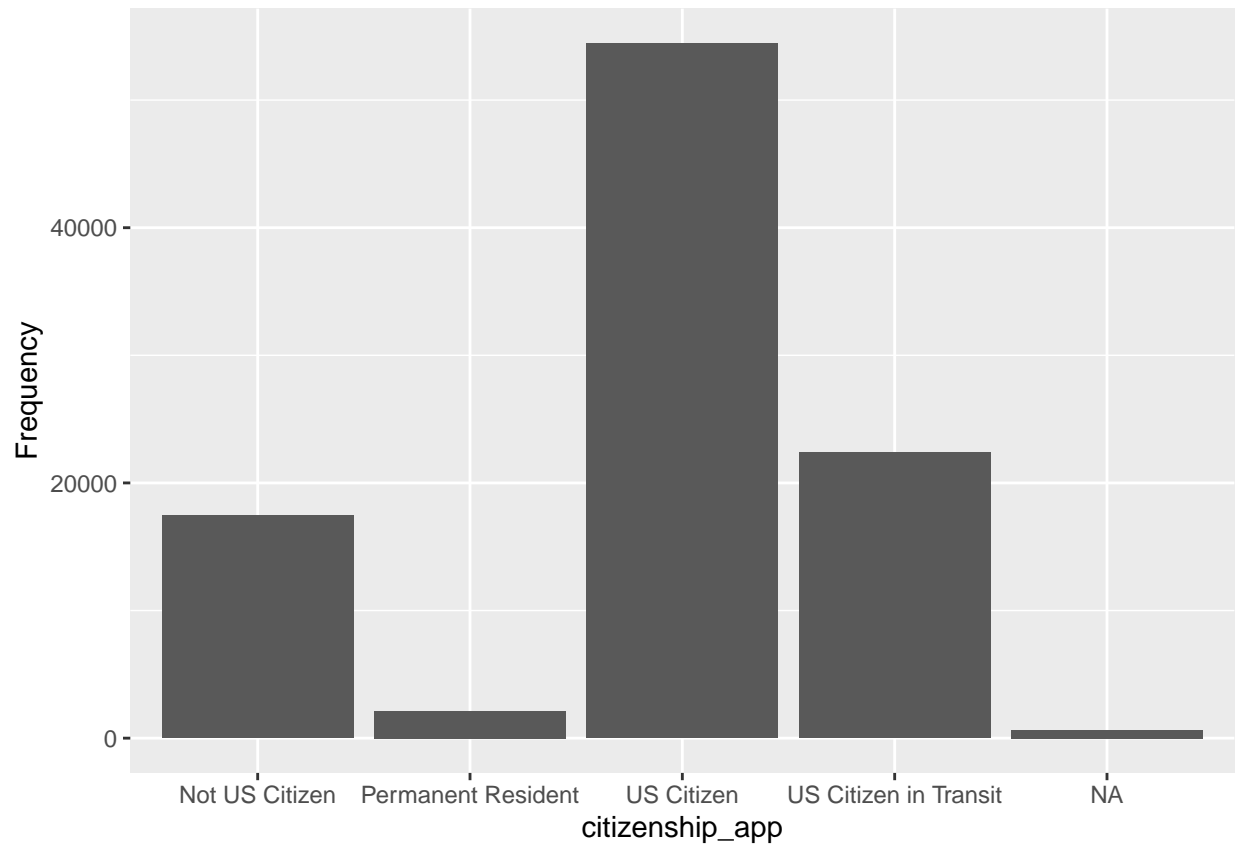


```
[1] -----  
[1] Variable: urm, type: numeric, requested: yes  
[1] Values (3 unique): 0, NA, 1  
[1] Missing: 5.8%  
[1] Most missing: 2008 8.2%, Least missing: 2012 4.3%
```

Warning: Removed 5582 rows containing non-finite values ('stat_count()').

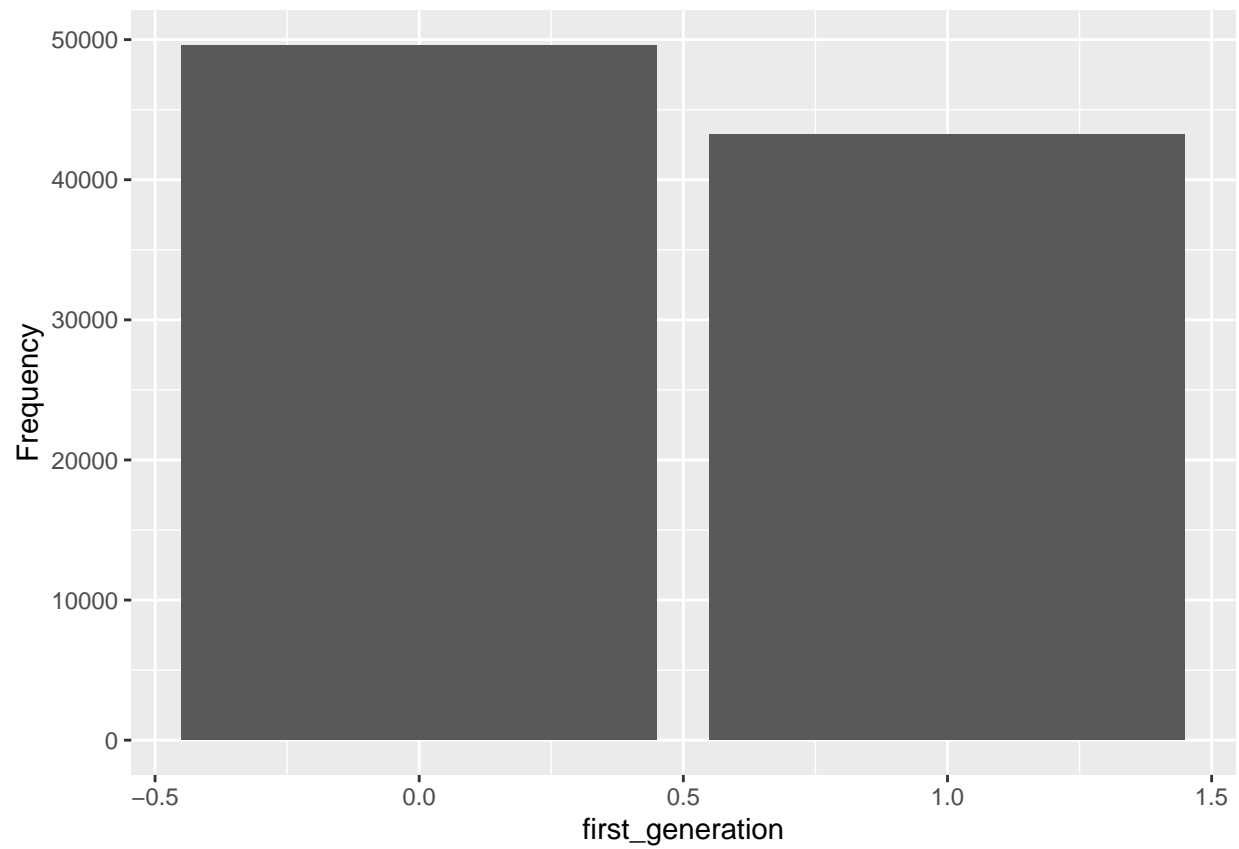


```
[1] -----  
[1] Variable: citizenship_app, type: character, requested: yes  
[1] Values (5 unique): US Citizen in Transit, Not US Citizen, Permanent Resident, NA, US Citizen  
[1] Missing: 0.7%
```

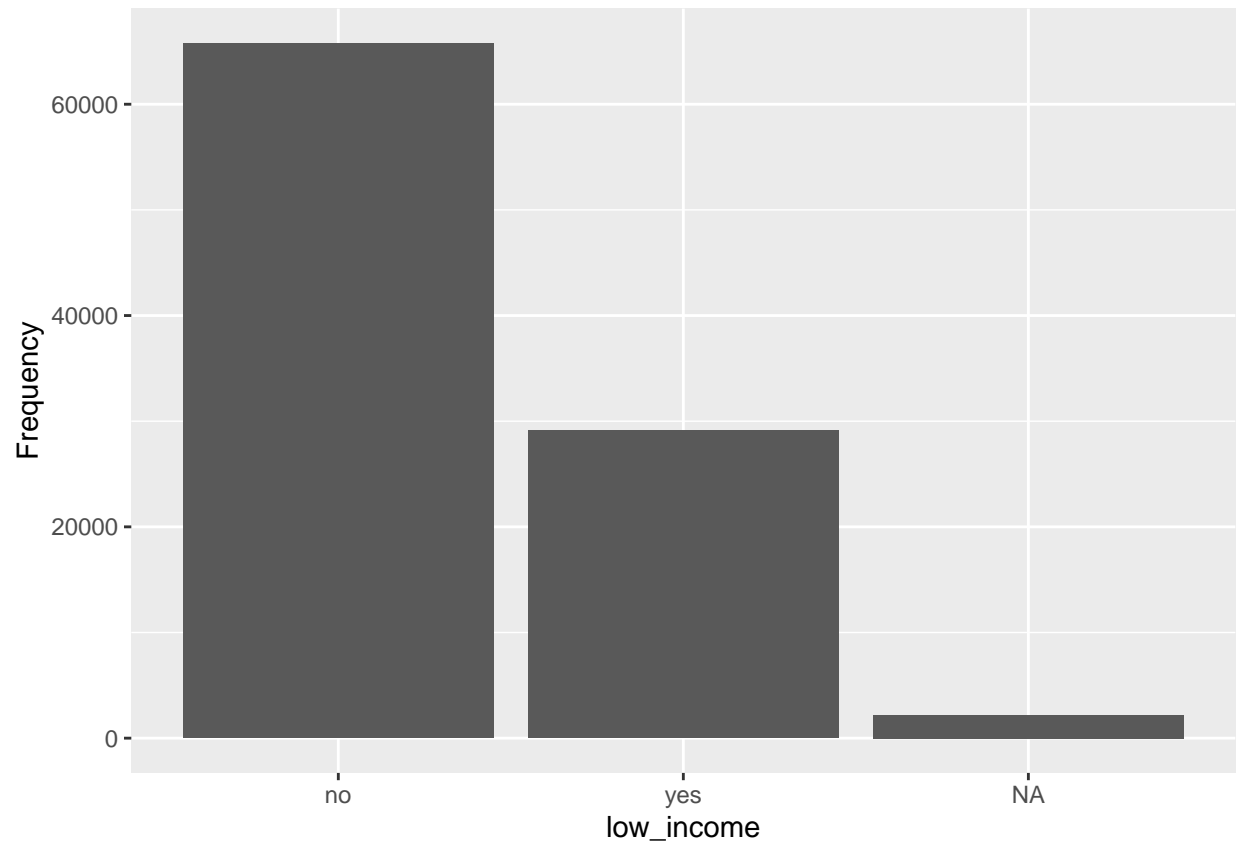


```
[1] -----  
[1] Variable: first_generation, type: numeric, requested: yes  
[1] Values (3 unique): 1, NA, 0  
[1] Missing: 4.3%  
[1] Most missing: 2006 8.3%, Least missing: 2017 1.9%
```

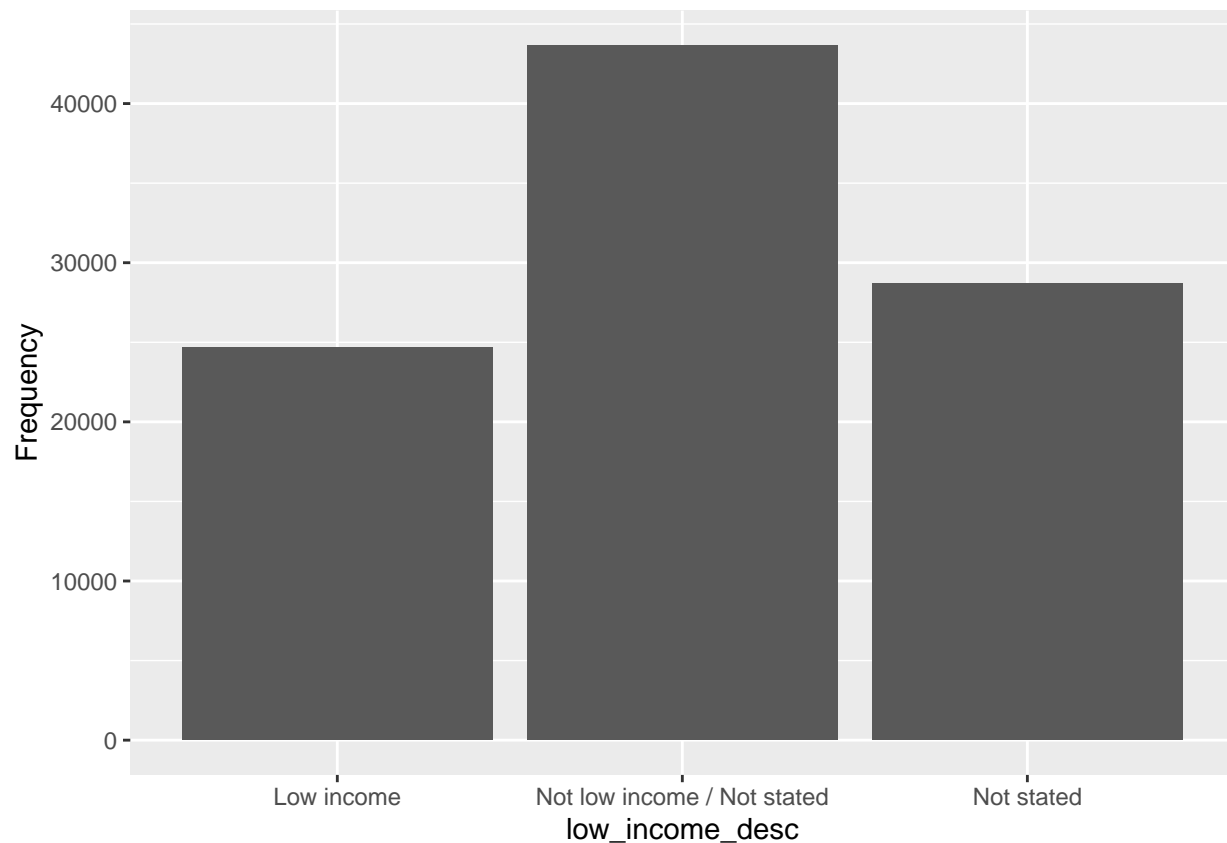
Warning: Removed 4221 rows containing non-finite values ('stat_count()').



```
[1] -----  
[1] Variable: low_income, type: character, requested: yes  
[1] Values (3 unique): no, NA, yes  
[1] Missing: 2.3%  
[1] Most missing: 2015 3.2%, Least missing: 2006 0.3%
```

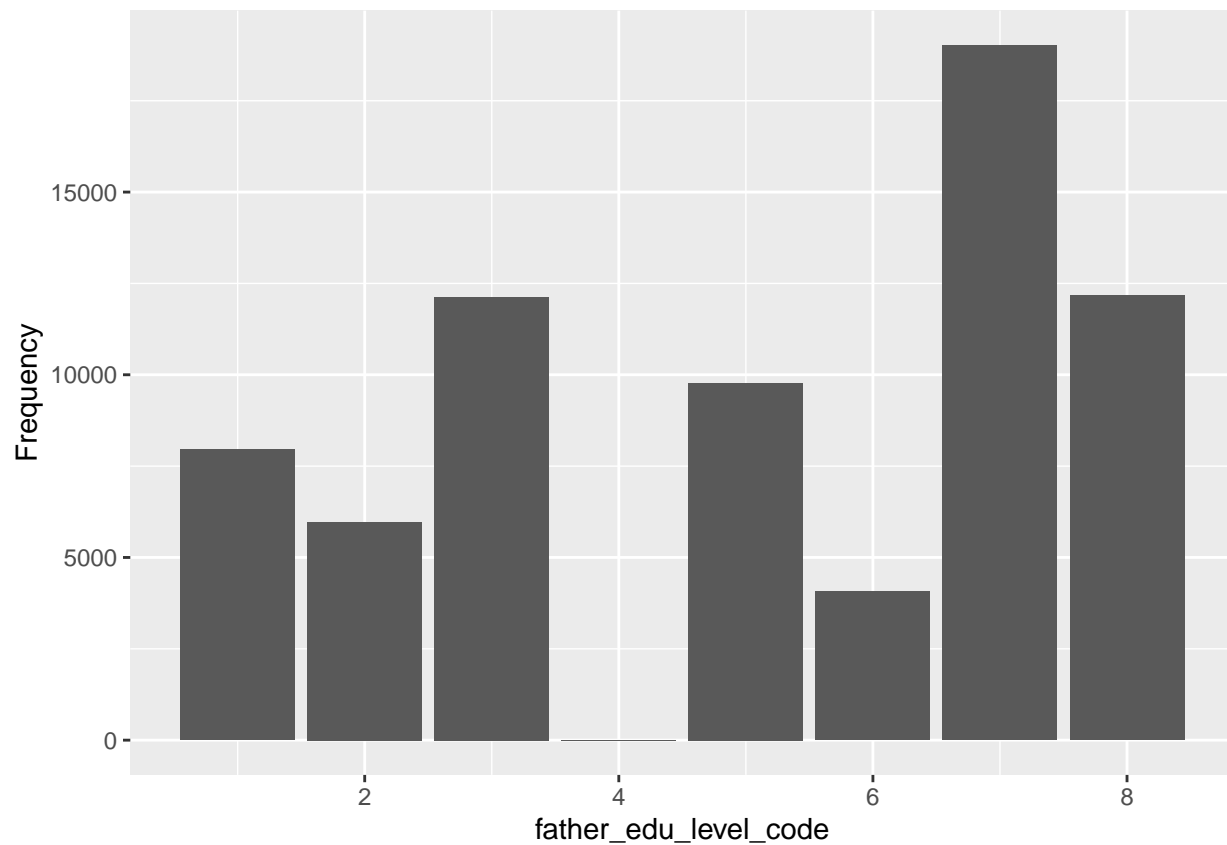


```
[1] -----  
[1] Variable: low_income_desc, type: character, requested: yes  
[1] Values (3 unique): Not low income / Not stated, Not stated, Low income  
[1] Missing: 0%
```

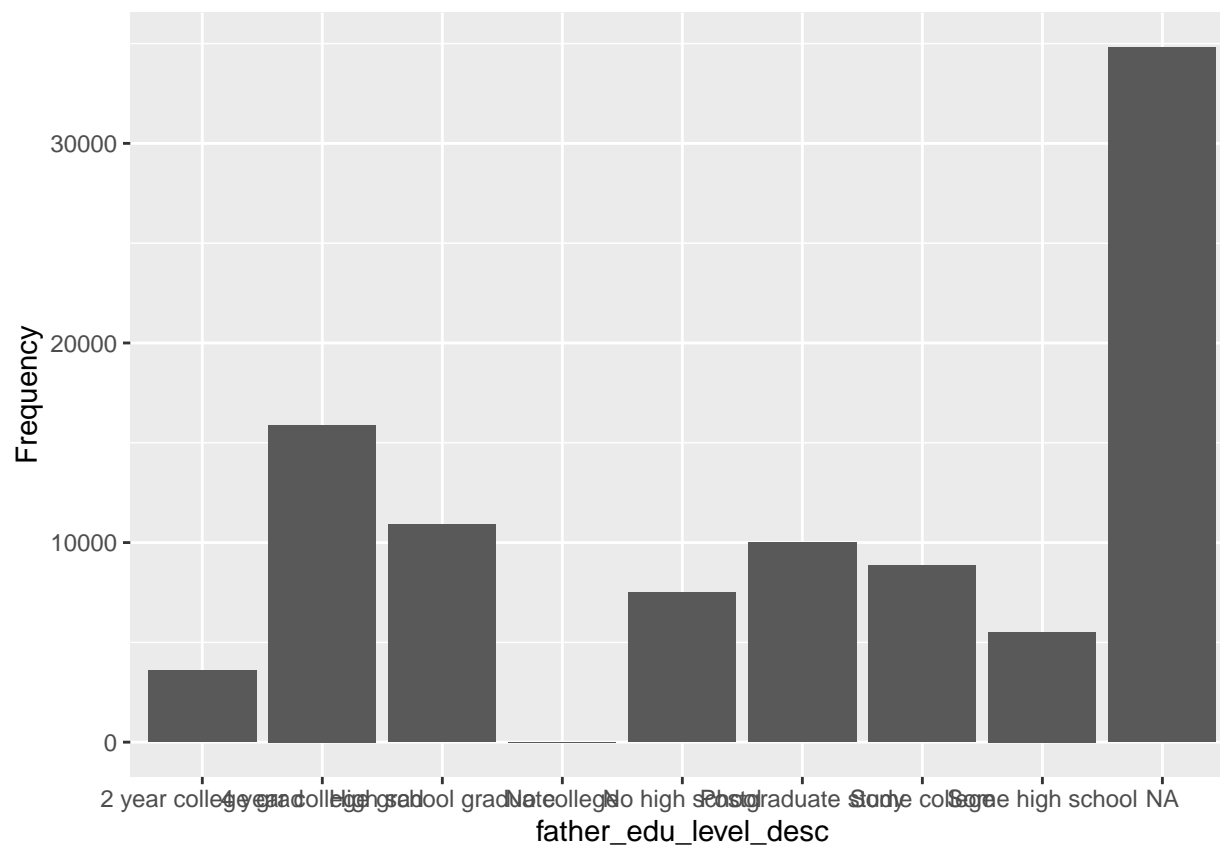


```
[1] should not be used as a predictor
[1] -----
[1] Variable: father_edu_level_code, type: numeric, requested: yes
[1] Values (9 unique): 3, NA, 8, 1, 5, ...
[1] Missing: 26.8%
Group.1 father_edu_level_code
1      2006      0.94870588
2      2007      0.84244657
3      2008      0.56529765
4      2009      0.45494210
5      2010      0.12548292
6      2011      0.10169972
7      2012      0.09752945
8      2013      0.10401364
9      2014      0.11367188
10     2015      0.11560767
11     2016      0.11077619
12     2017      0.10497067
13     2018      0.11226133
```

Warning: Removed 25967 rows containing non-finite values ('stat_count()').

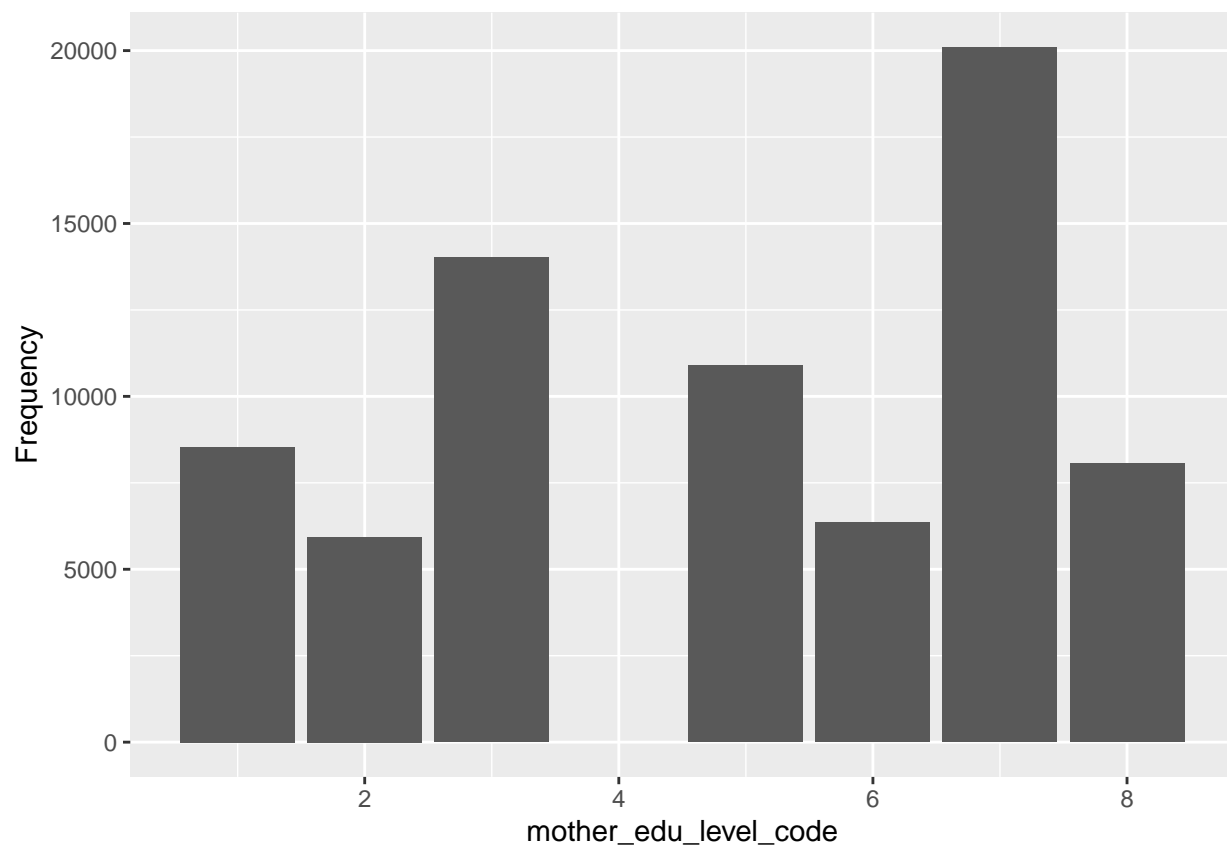


```
[1] -----
[1] Variable: father_educ_level_desc, type: character, requested: yes
[1] Values (9 unique): High school graduate, NA, Postgraduate study, No high school, Some college, ...
[1] Missing: 35.9%
Group.1 father_educ_level_desc
1      2006      0.9487059
2      2007      0.8424466
3      2008      0.5652976
4      2009      0.4549421
5      2010      0.3039716
6      2011      0.2092068
7      2012      0.2392991
8      2013      0.2594439
9      2014      0.2670573
10     2015      0.2206253
11     2016      0.3290989
12     2017      0.1710404
13     2018      0.1212330
```

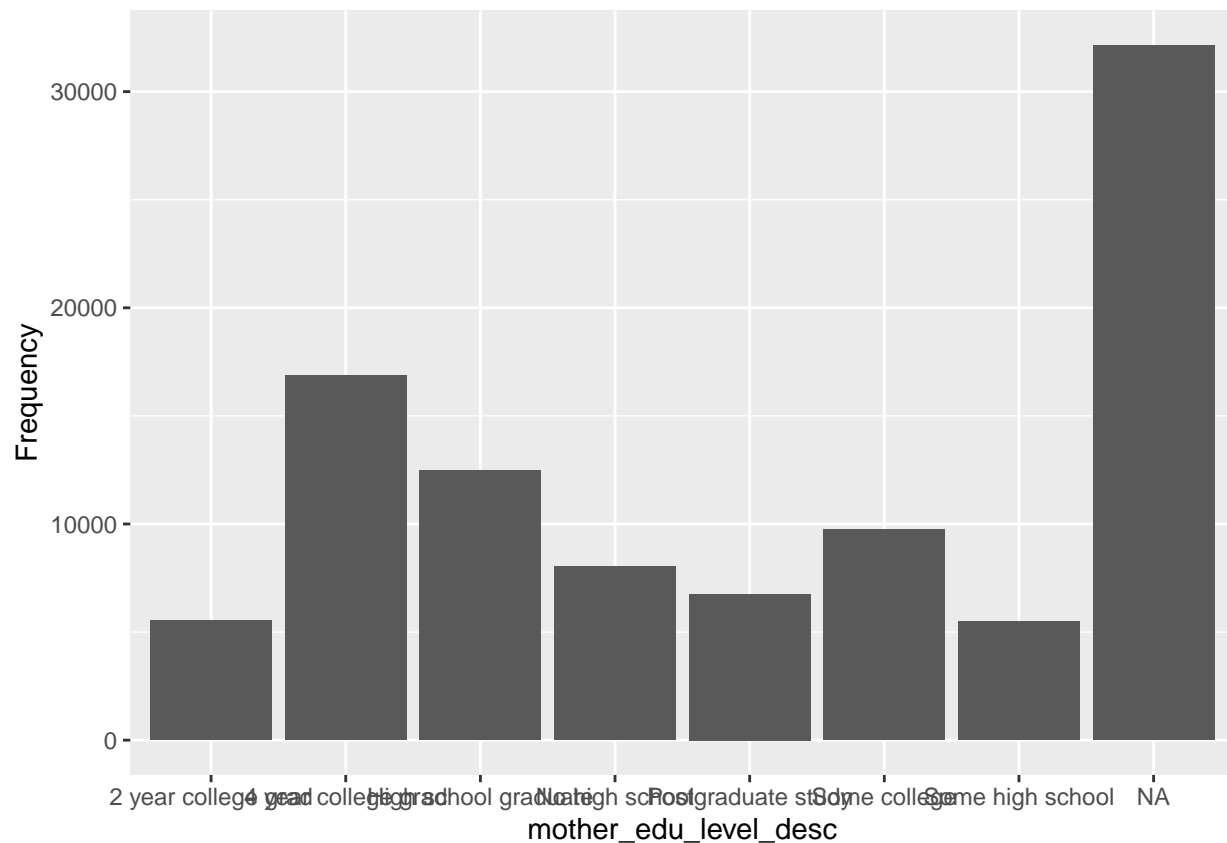



```
[1] should not be used as a predictor
[1] -----
[1] Variable: mother_edu_level_code, type: numeric, requested: yes
[1] Values (8 unique): 6, NA, 7, 1, 8, ...
[1] Missing: 23.9%
Group.1 mother_edu_level_code
1      2006      0.94729412
2      2007      0.83537214
3      2008      0.55006922
4      2009      0.43094479
5      2010      0.09967548
6      2011      0.06770538
7      2012      0.06190750
8      2013      0.07109129
9      2014      0.07630208
10     2015      0.07778618
11     2016      0.07632684
12     2017      0.06781929
13     2018      0.07522429
```

Warning: Removed 23172 rows containing non-finite values ('stat_count()').



```
[1] -----
[1] Variable: mother_educ_level_desc, type: character, requested: yes
[1] Values (8 unique): 2 year college grad, NA, 4 year college grad, No high school, Postgraduate study
[1] Missing: 33.1%
Group.1 mother_educ_level_desc
1      2006      0.94729412
2      2007      0.83537214
3      2008      0.55006922
4      2009      0.43094479
5      2010      0.28017308
6      2011      0.17464589
7      2012      0.20338983
8      2013      0.22547219
9      2014      0.22929688
10     2015      0.18330812
11     2016      0.30487674
12     2017      0.13666770
13     2018      0.08442604
```



[1] should not be used as a predictor

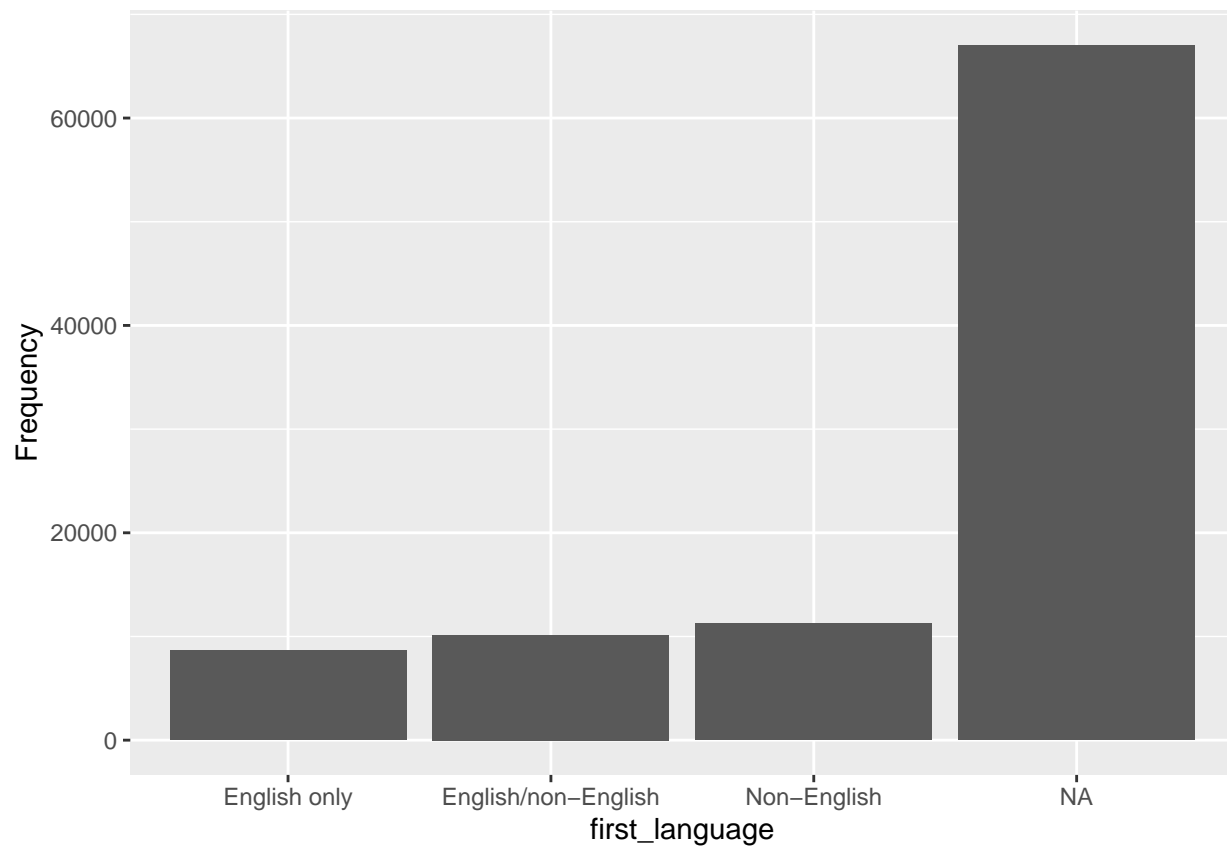
[1] -----

[1] Variable: first_language, type: character, requested: yes

[1] Values (4 unique): English only, NA, Non-English, English/non-English

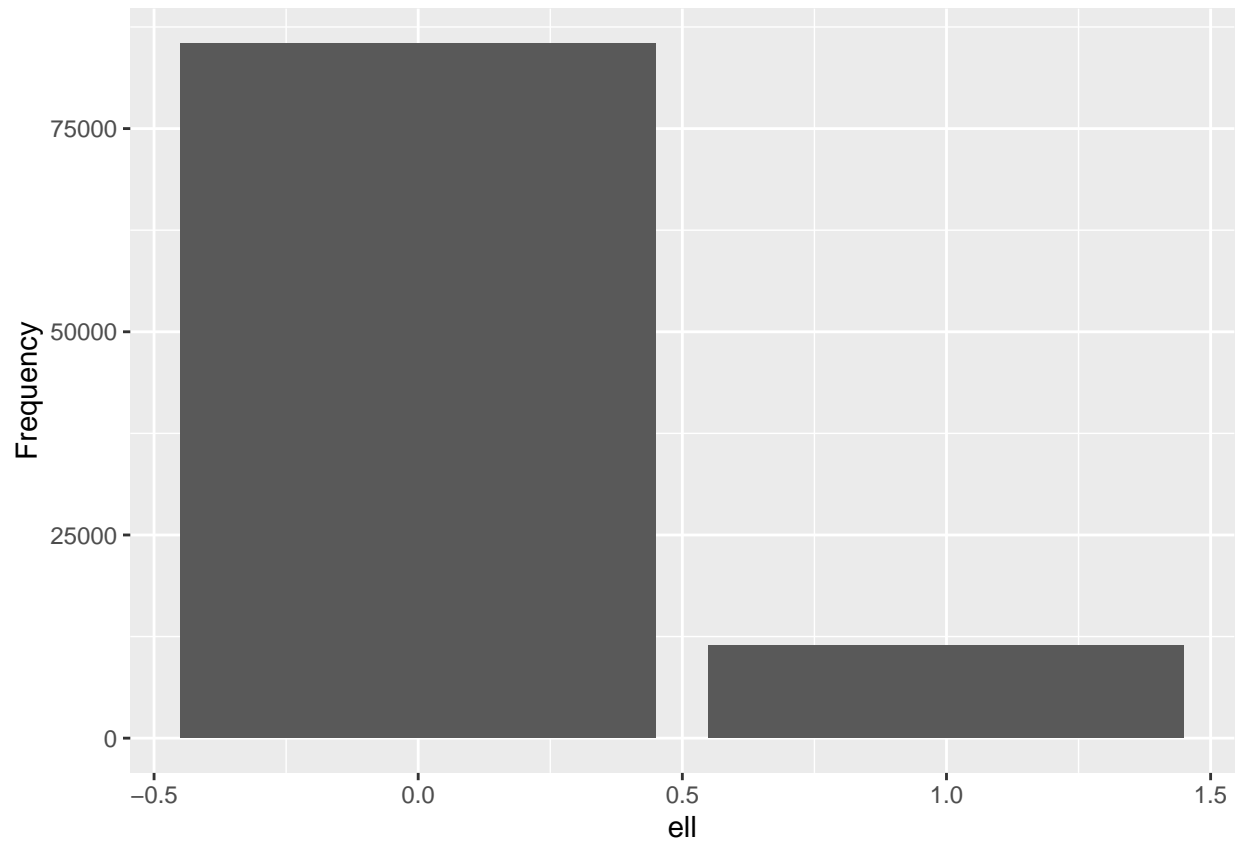
[1] Missing: 69.1%

Group.1 first_language		
1	2006	0.99701961
2	2007	0.99749447
3	2008	0.99630826
4	2009	0.99513341
5	2010	0.99613661
6	2011	0.99079320
7	2012	0.98850905
8	2013	0.97520986
9	2014	0.88776042
10	2015	0.41805345
11	2016	0.27322640
12	2017	0.08984254
13	2018	0.02058891

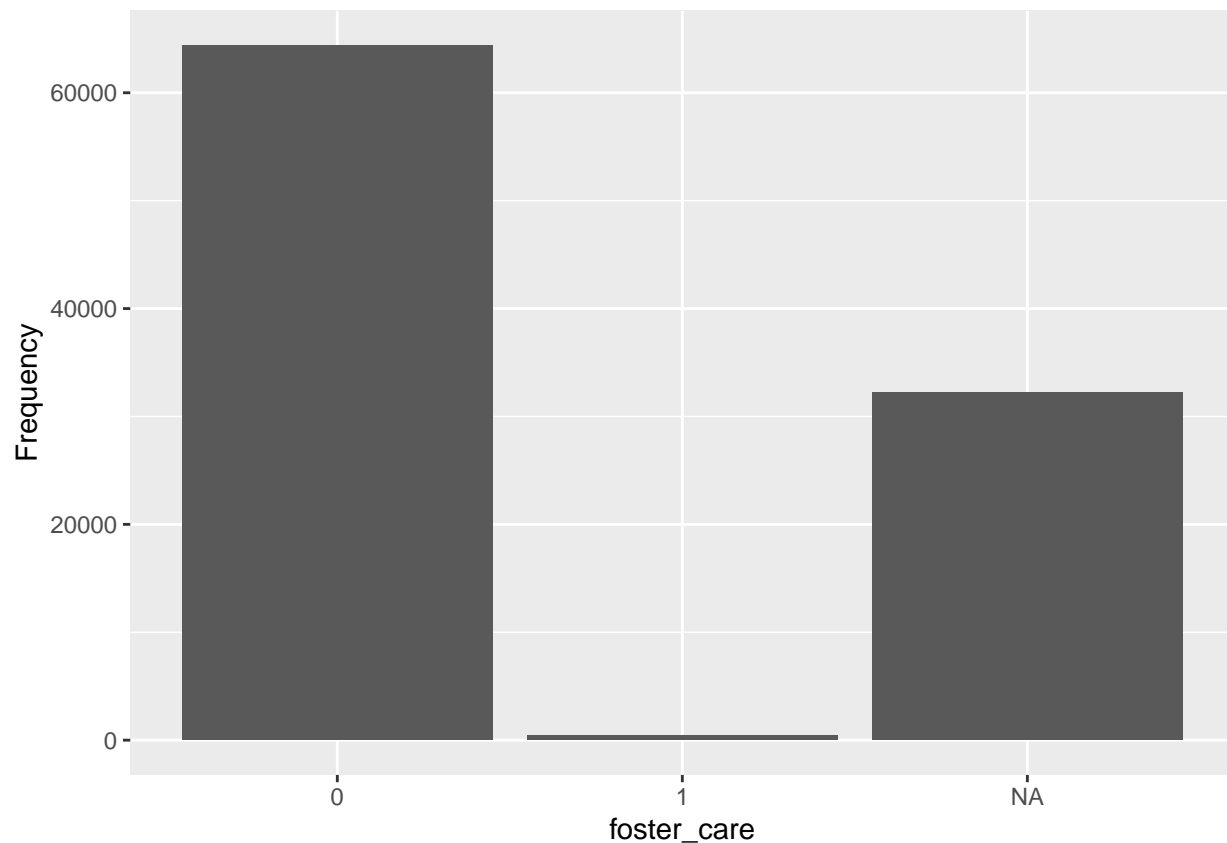


```
[1] -----  
[1] Variable: ell, type: numeric, requested: yes  
[1] Values (3 unique): 0, NA, 1  
[1] Missing: 0.2%
```

Warning: Removed 189 rows containing non-finite values ('stat_count()').

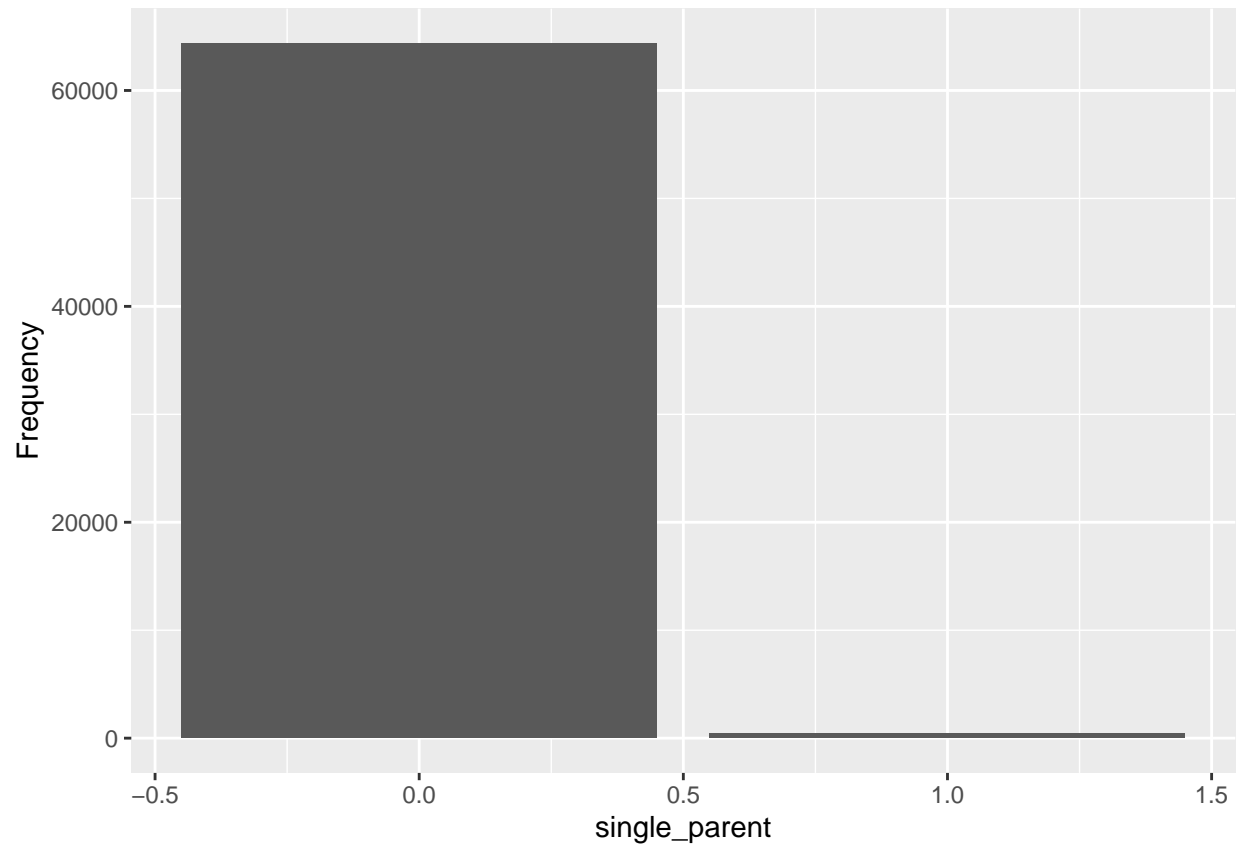


```
[1] -----
[1] Variable: foster_care, type: character, requested: yes
[1] Values (3 unique): 0, NA, 1
[1] Missing: 33.2%
  Group.1 foster_care
1    2006 0.9989019608
2    2007 0.9986735446
3    2008 0.9918474081
4    2009 0.9927840242
5    2010 0.9932004327
6    2011 0.0000000000
7    2012 0.0002872738
8    2013 0.0001311647
9    2014 0.0036458333
10   2015 0.0031517902
11   2016 0.0047367854
12   2017 0.0015436863
13   2018 0.0205889119
```

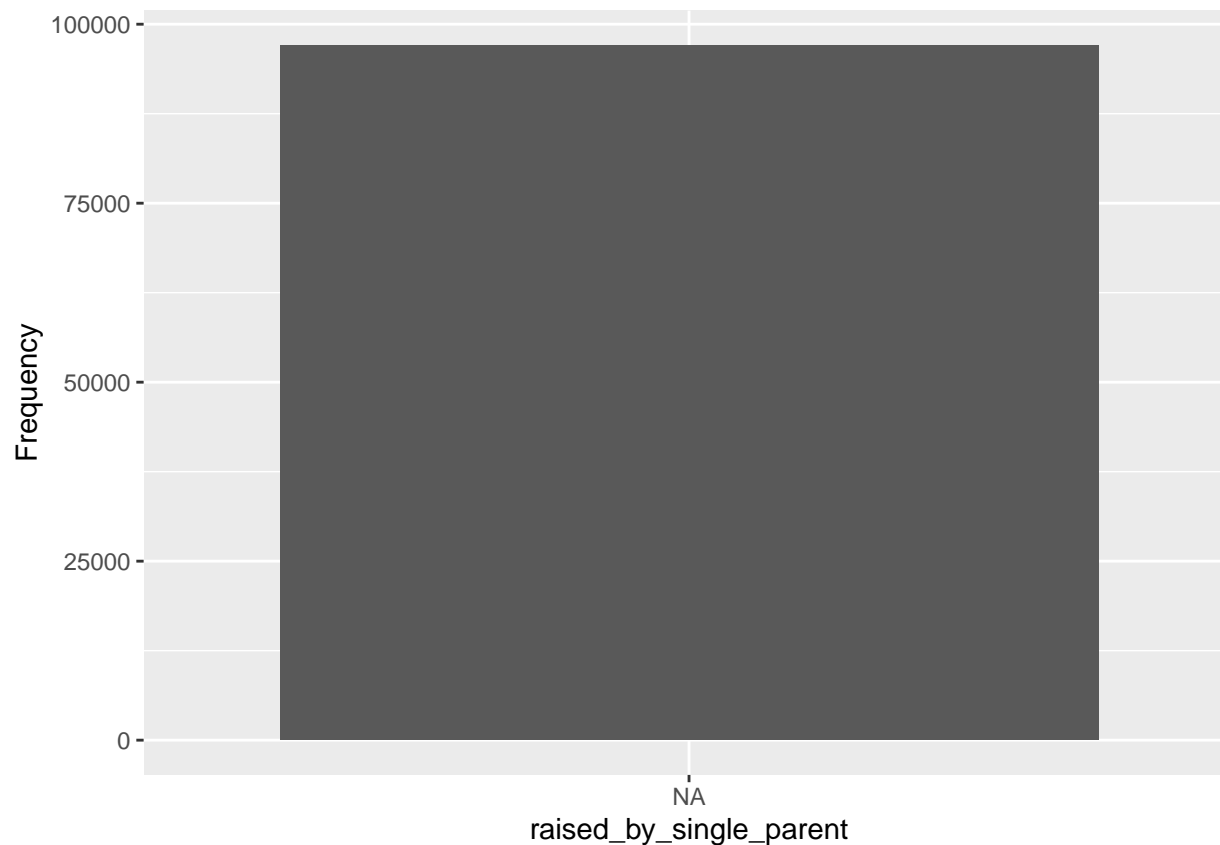


```
[1] -----
[1] Variable: single_parent, type: numeric, requested: yes
[1] Values (3 unique): 0, NA, 1
[1] Missing: 33.2%
Group.1 single_parent
1      2006  0.9990588235
2      2007  0.9985261606
3      2008  0.9955391478
4      2009  0.9974827991
5      2010  0.9970638232
6      2011  0.0004249292
7      2012  0.0005745475
8      2013  0.0003934942
9      2014  0.0037760417
10     2015  0.0037821483
11     2016  0.0055980192
12     2017  0.0028815478
13     2018  0.0113871636
```

Warning: Removed 32260 rows containing non-finite values ('stat_count()').

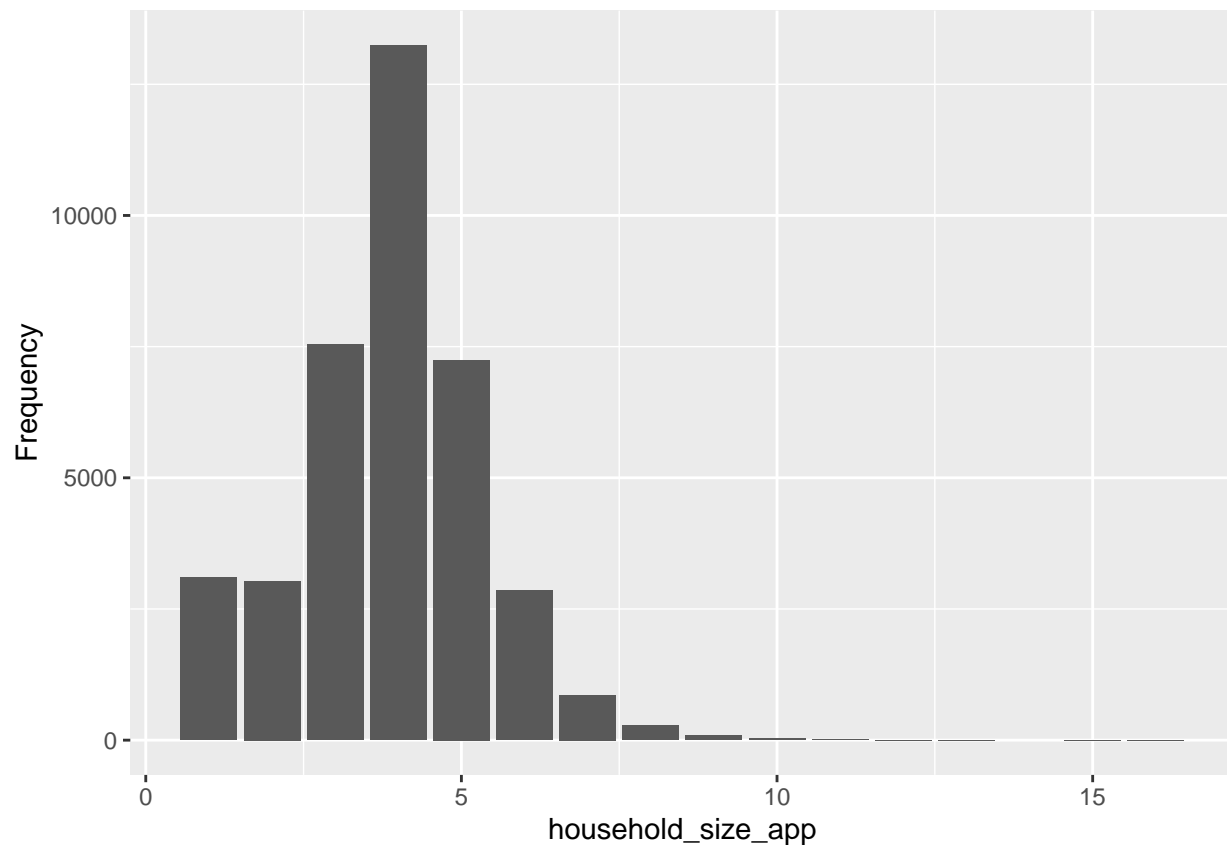


```
[1] -----  
[1] Variable: raised_by_single_parent, type: logical, requested: yes  
[1] Values (1 unique): NA  
[1] Missing: 100%
```

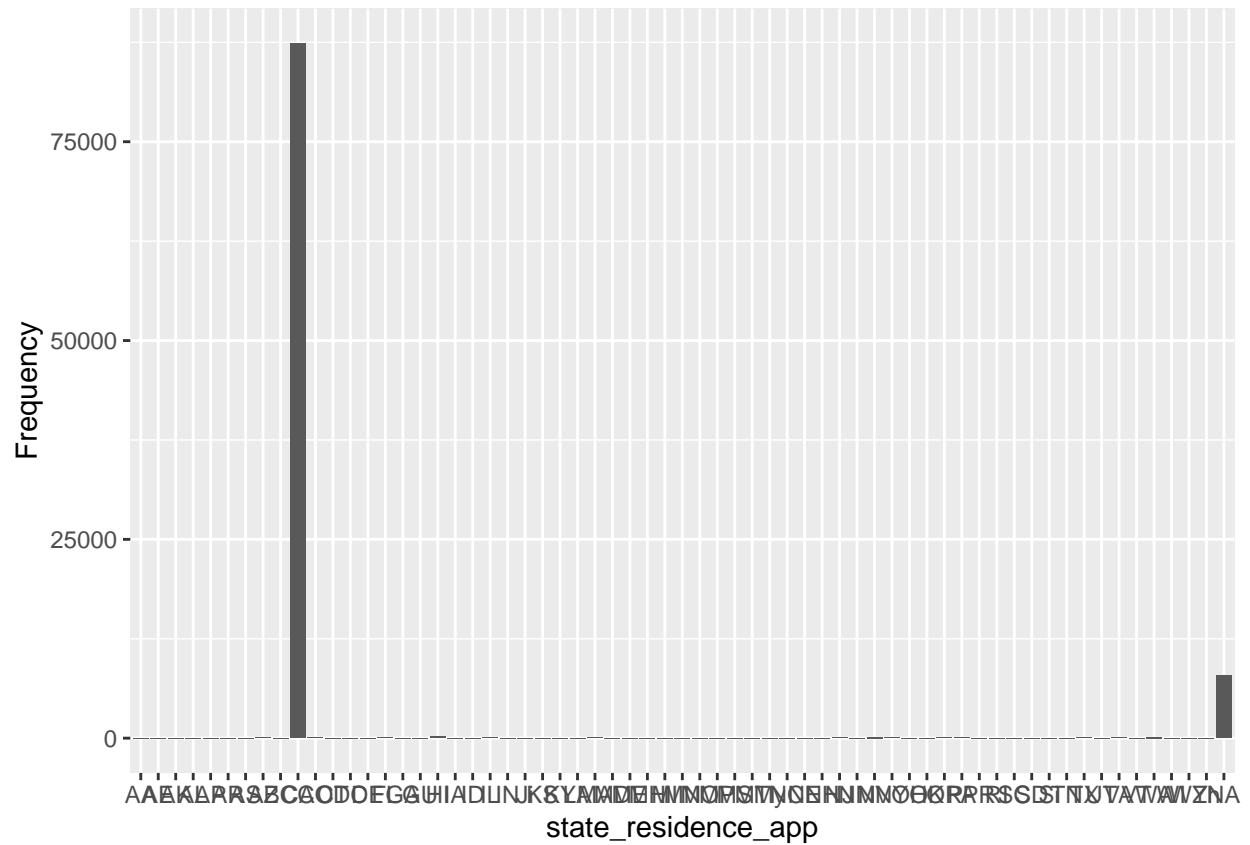


```
[1] -----
[1] Variable: household_size_app, type: numeric, requested: yes
[1] Values (16 unique): NA, 5, 4, 1, 6, ...
[1] Missing: 60.5%
Group.1 household_size_app
1      2006      0.9460392
2      2007      0.8266765
3      2008      0.5234579
4      2009      0.3965430
5      2010      0.2352032
6      2011      0.1507082
7      2012      0.1844298
8      2013      0.2134050
9      2014      0.3078125
10     2015      0.7289460
11     2016      0.9992464
12     2017      1.0000000
13     2018      1.0000000
```

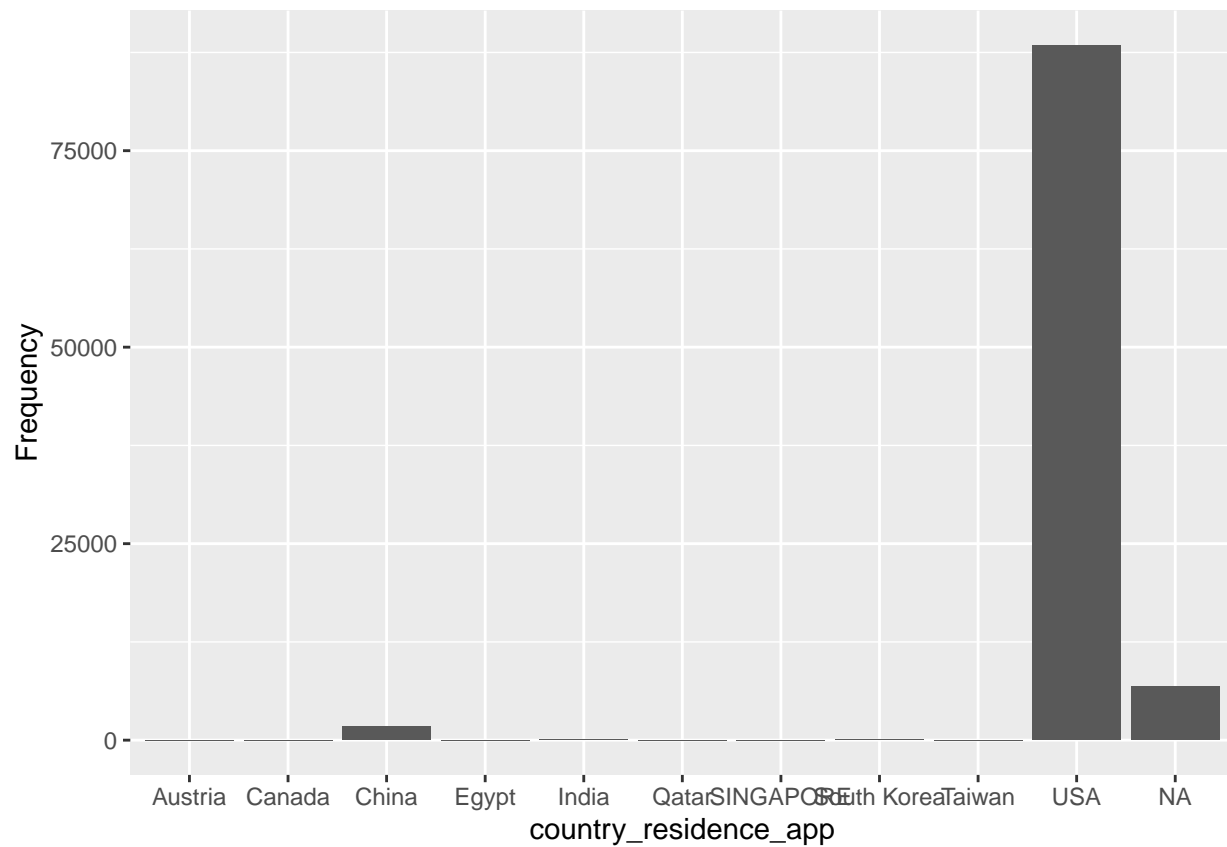
Warning: Removed 58742 rows containing non-finite values ('stat_count()').



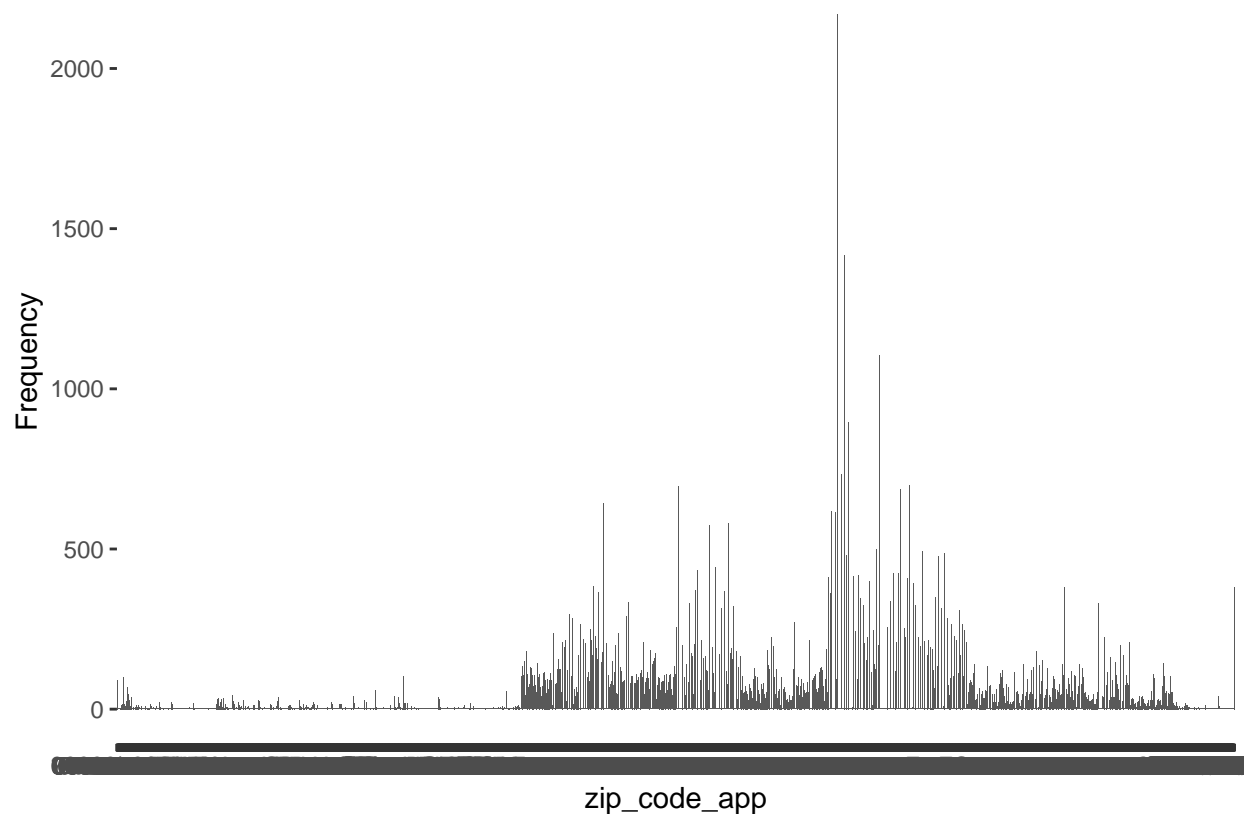
```
[1] -----
[1] Variable: city_residence_app, type: character, requested: yes
[1] Values (5040 unique): WOODLAND HILLS, NA, CORONA, INDIO, SAN JOSE, ...
[1] Missing: 0.2%
[1] -----
[1] Variable: state_residence_app, type: character, requested: yes
[1] Values (63 unique): CA, NA, HI, WA, PR, ...
[1] Missing: 8.1%
[1] Most missing: 2016 14.5%, Least missing: 2006 1.3%
```



```
[1] -----
[1] Variable: country_residence_app, type: character, requested: yes
[1] Values (11 unique): USA, NA, South Korea, China, SINGAPORE, ...
[1] Missing: 7%
[1] Most missing: 2015 10.5%, Least missing: 2006 2.1%
```

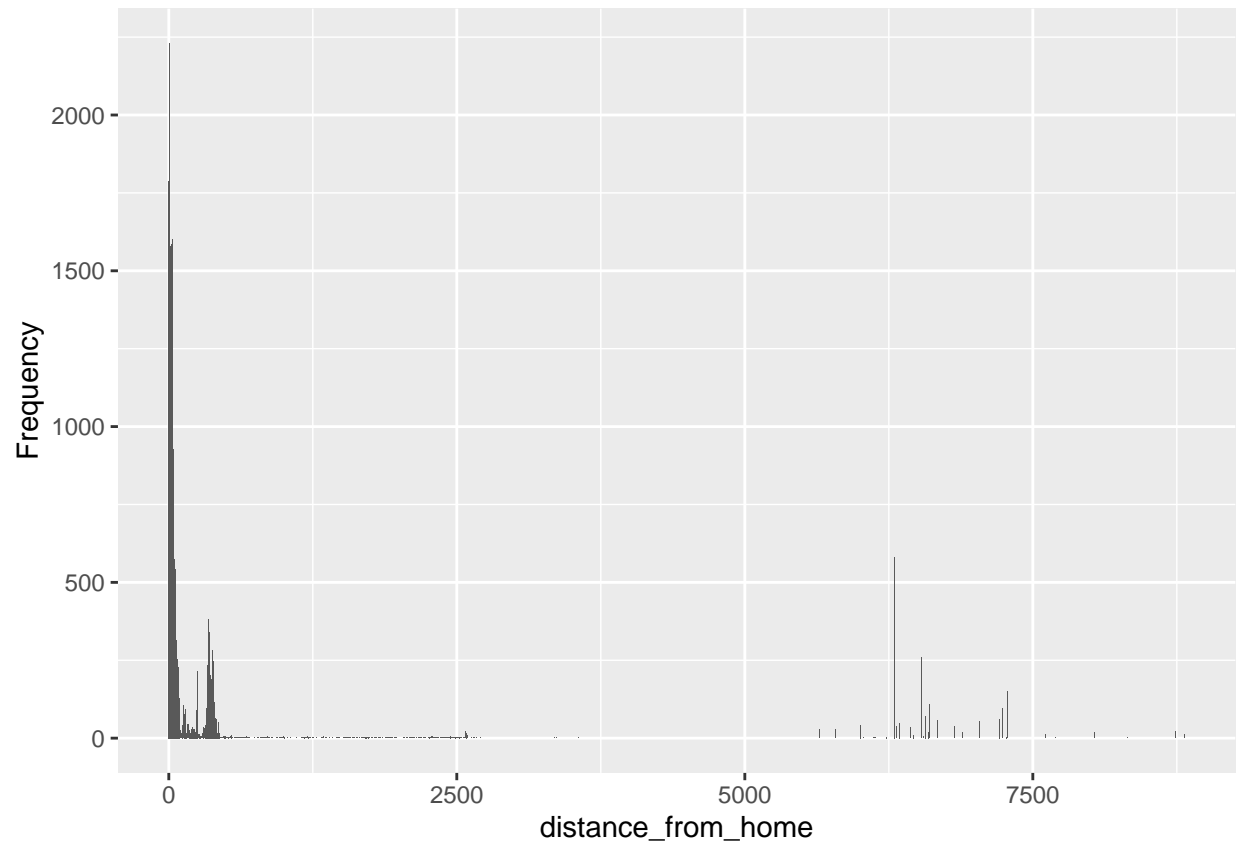


```
[1] -----
[1] Variable: zip_code_app, type: character, requested: yes
[1] Values (11519 unique): 91364, NA, 92881, 92201, 95120-3406, ...
[1] Missing: 0.4%
```

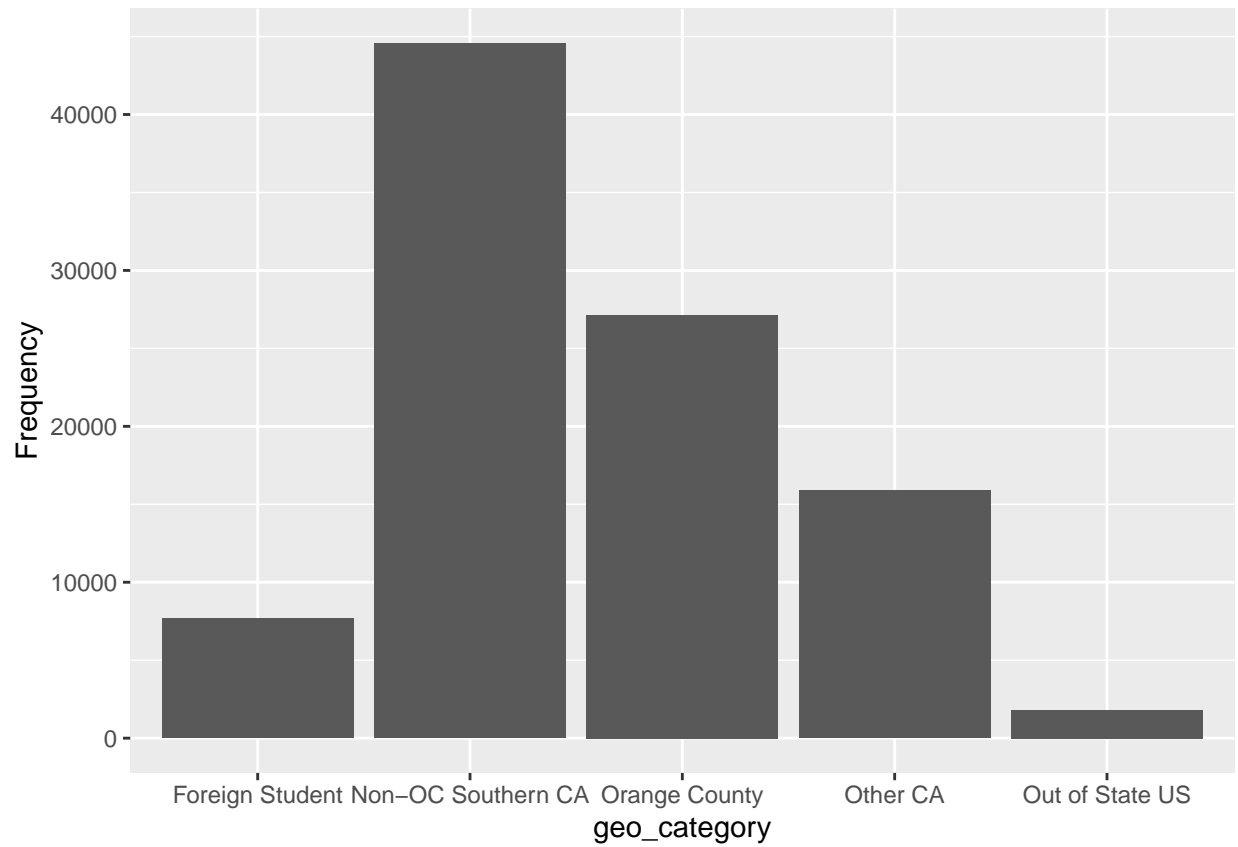


```
[1] -----
[1] Variable: distance_from_home, type: numeric, requested: yes
[1] Values (859 unique): 56, NA, 6006, 6302, 6604, ...
[1] Missing: 34.2%
  Group.1 distance_from_home
1      2006      0.02054902
2      2007      0.04023581
3      2008      0.04568528
4      2009      0.05185434
5      2010      0.05532375
6      2011      0.07053824
7      2012      0.09796036
8      2013      0.12277020
9      2014      0.20286458
10     2015      0.60564801
11     2016      0.70330498
12     2017      0.88319440
13     2018      0.94870025
```

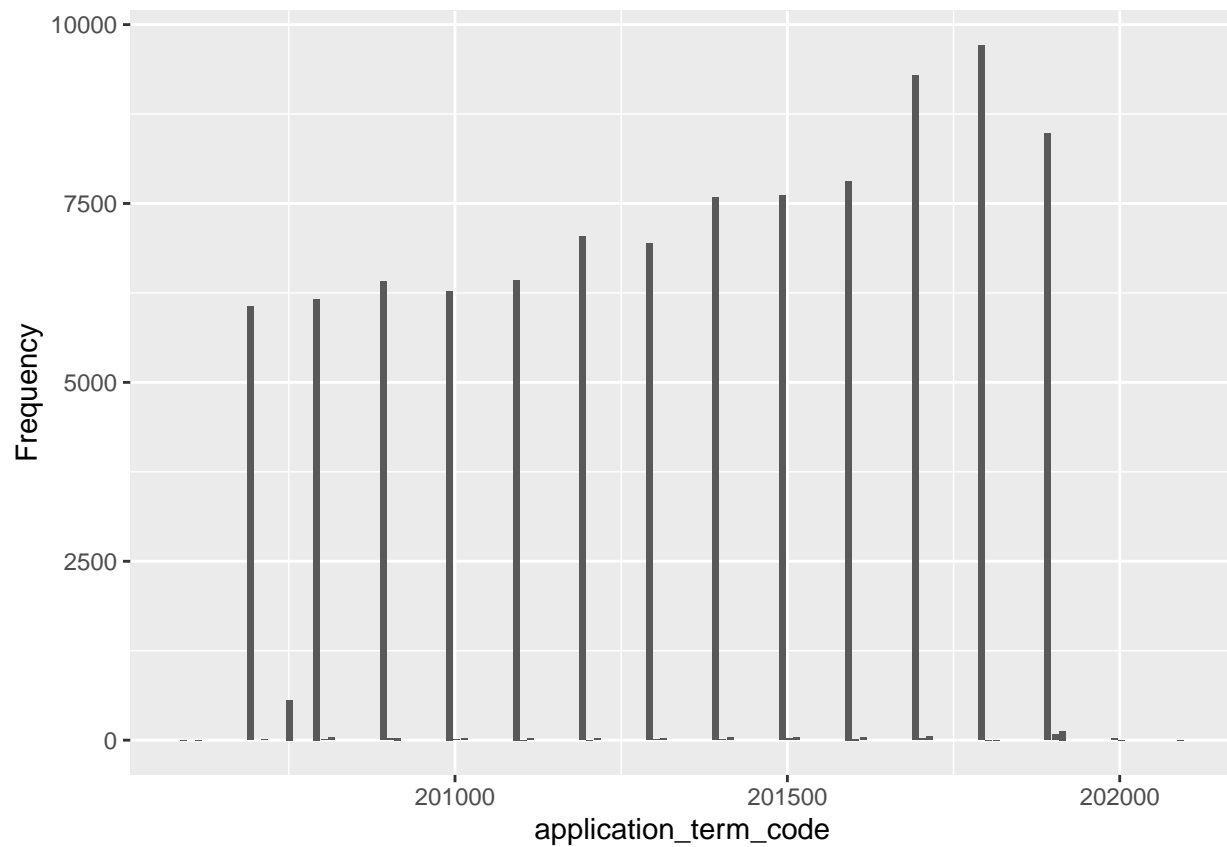
Warning: Removed 33209 rows containing non-finite values ('stat_count()').

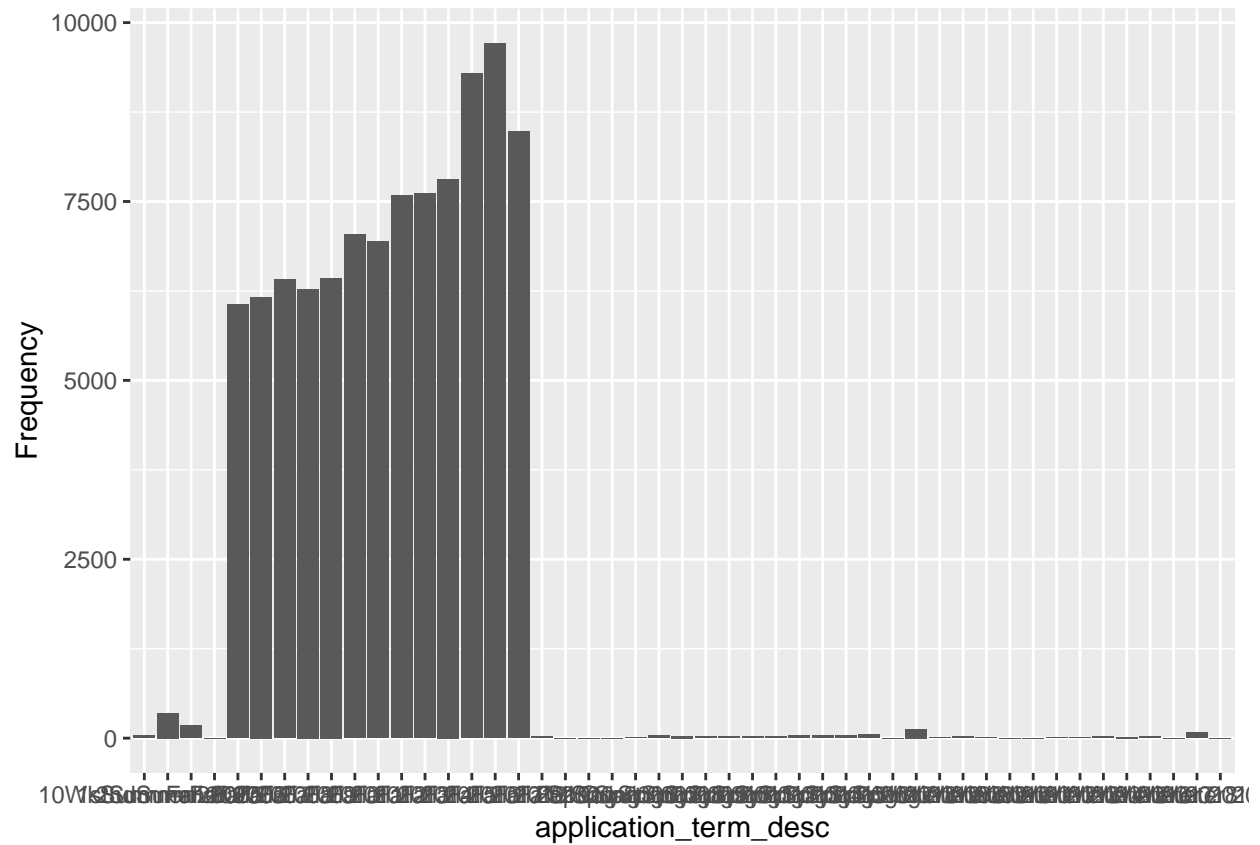


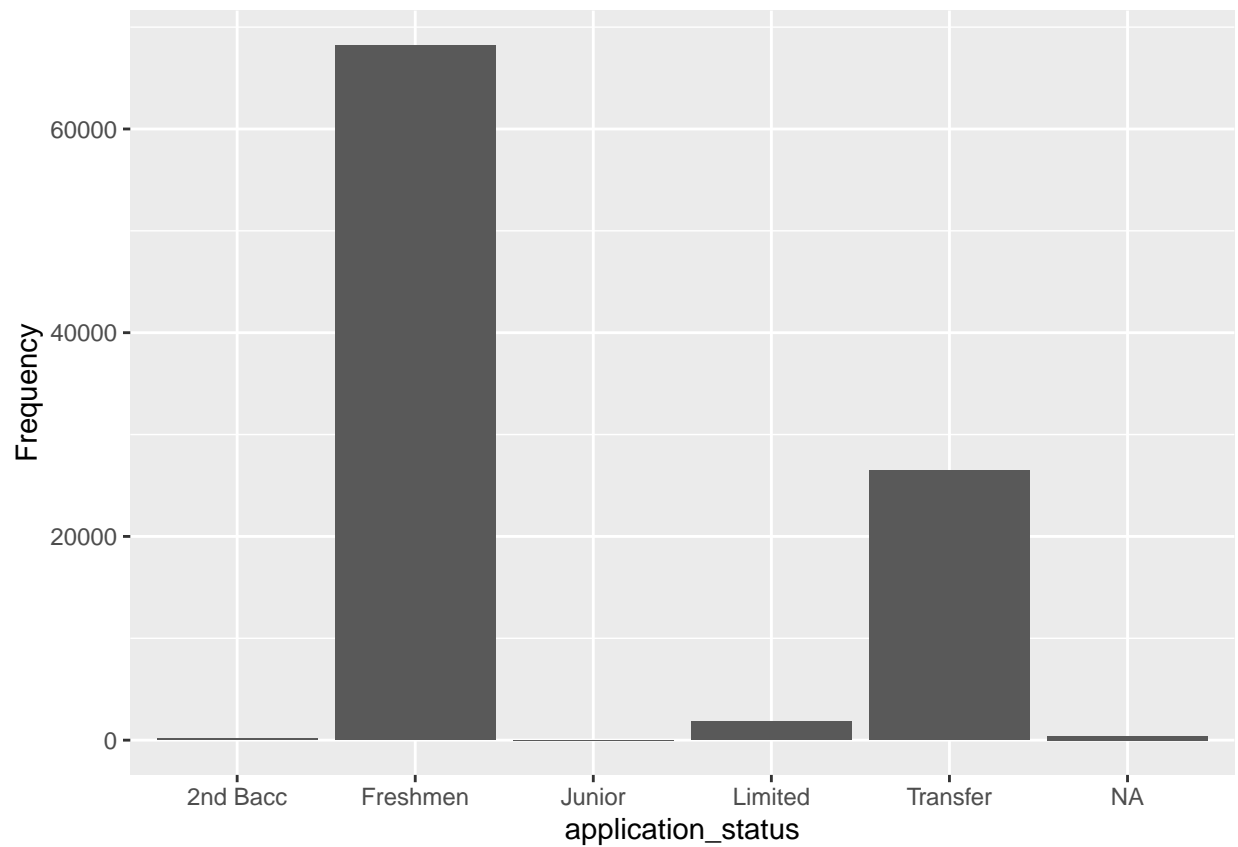
```
[1] -----  
[1] Variable: geo_category, type: character, requested: yes  
[1] Values (5 unique): Non-OC Southern CA, Other CA, Orange County, Out of State US, Foreign Student  
[1] Missing: 0%
```



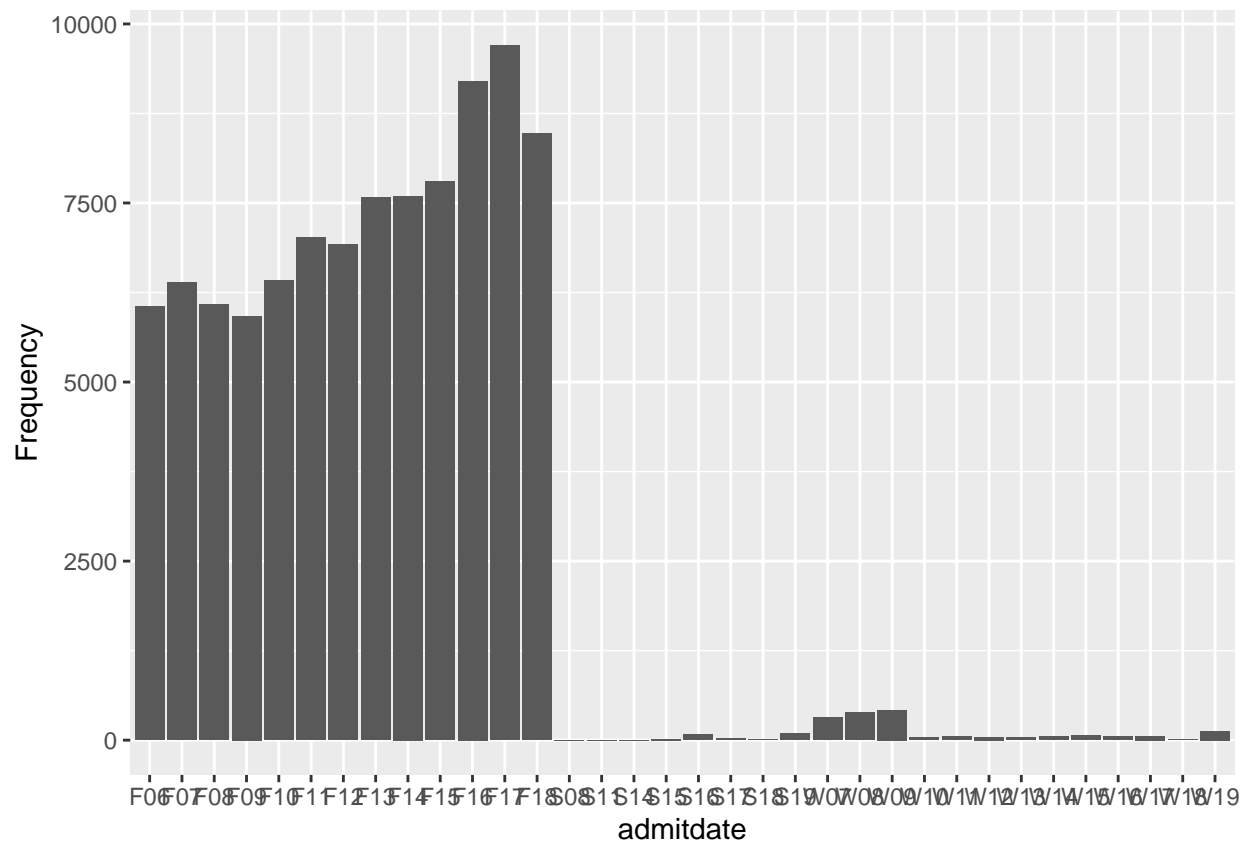
```
[1] -----  
[1] Variable: application_term_code, type: numeric, requested: yes  
[1] Values (44 unique): 202092, 201914, 201903, 201592, 201092, ...  
[1] Missing: 0%
```



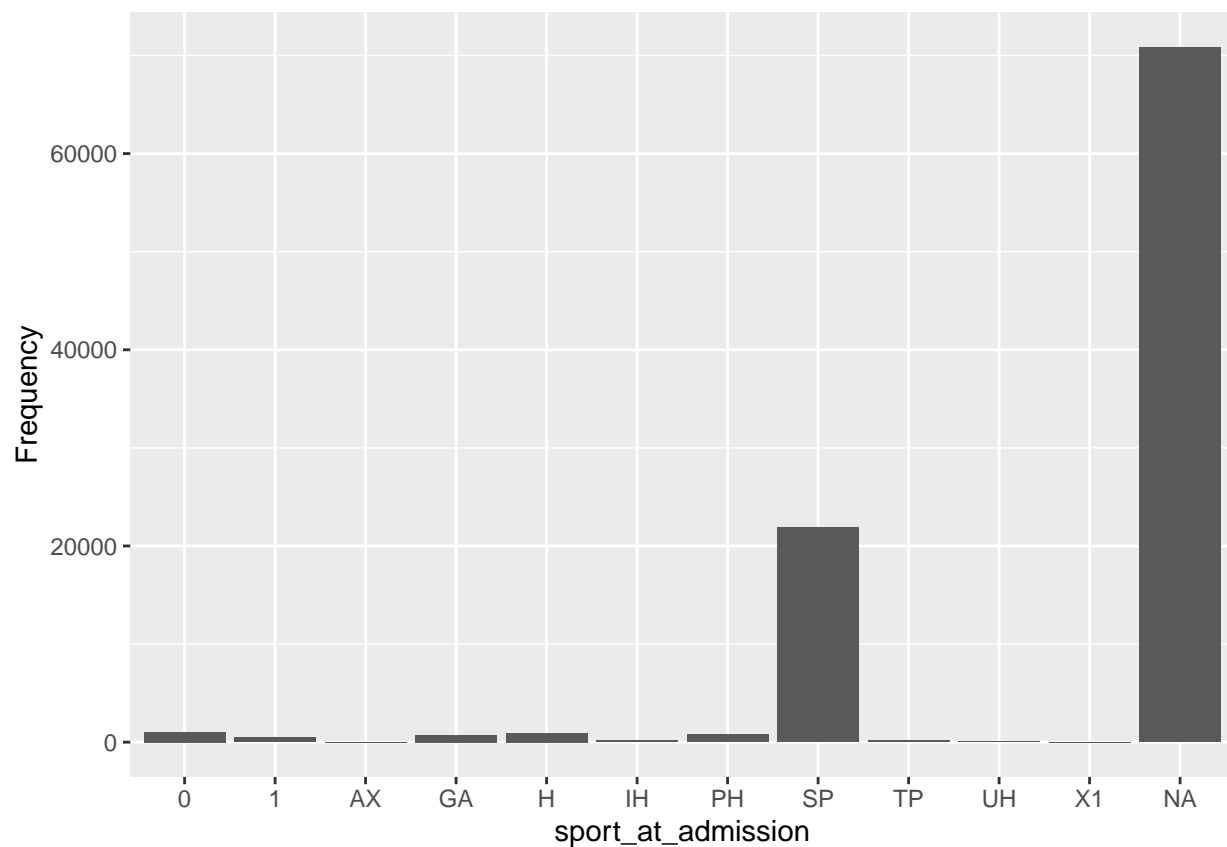




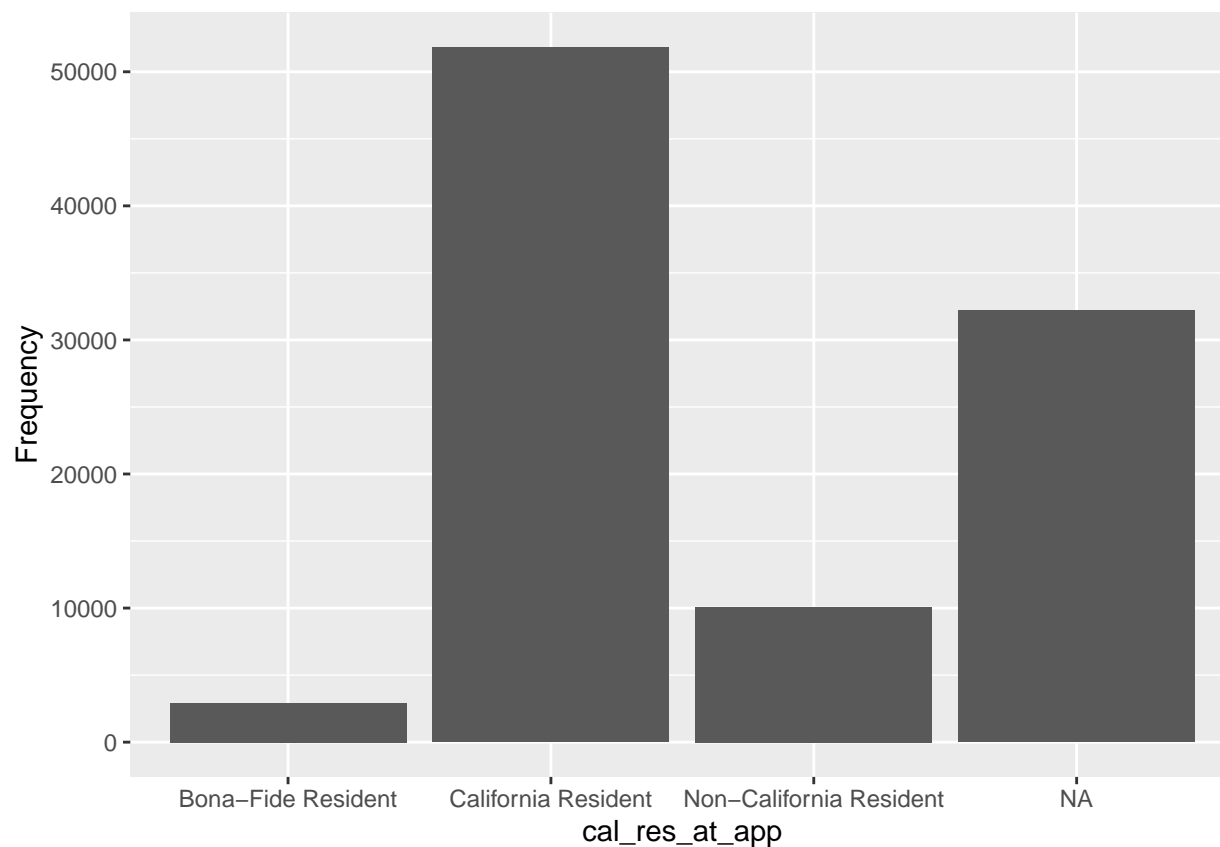
```
[1] -----  
[1] Variable: admitdate, type: character, requested: yes  
[1] Values (34 unique): S18, S19, W19, F15, F10, ...  
[1] Missing: 0%
```



```
[1] -----
[1] Variable: sport_at_admission, type: character, requested: yes
[1] Values (12 unique): NA, SP, GA, TP, 1, ...
[1] Missing: 73%
Group.1 sport_at_admission
1      2006      1.0000000
2      2007      1.0000000
3      2008      0.9993847
4      2009      1.0000000
5      2010      1.0000000
6      2011      0.1398017
7      2012      0.1875898
8      2013      0.2443599
9      2014      0.3363281
10     2015      0.7605900
11     2016      0.8873937
12     2017      0.9299166
13     2018      0.9998850
```

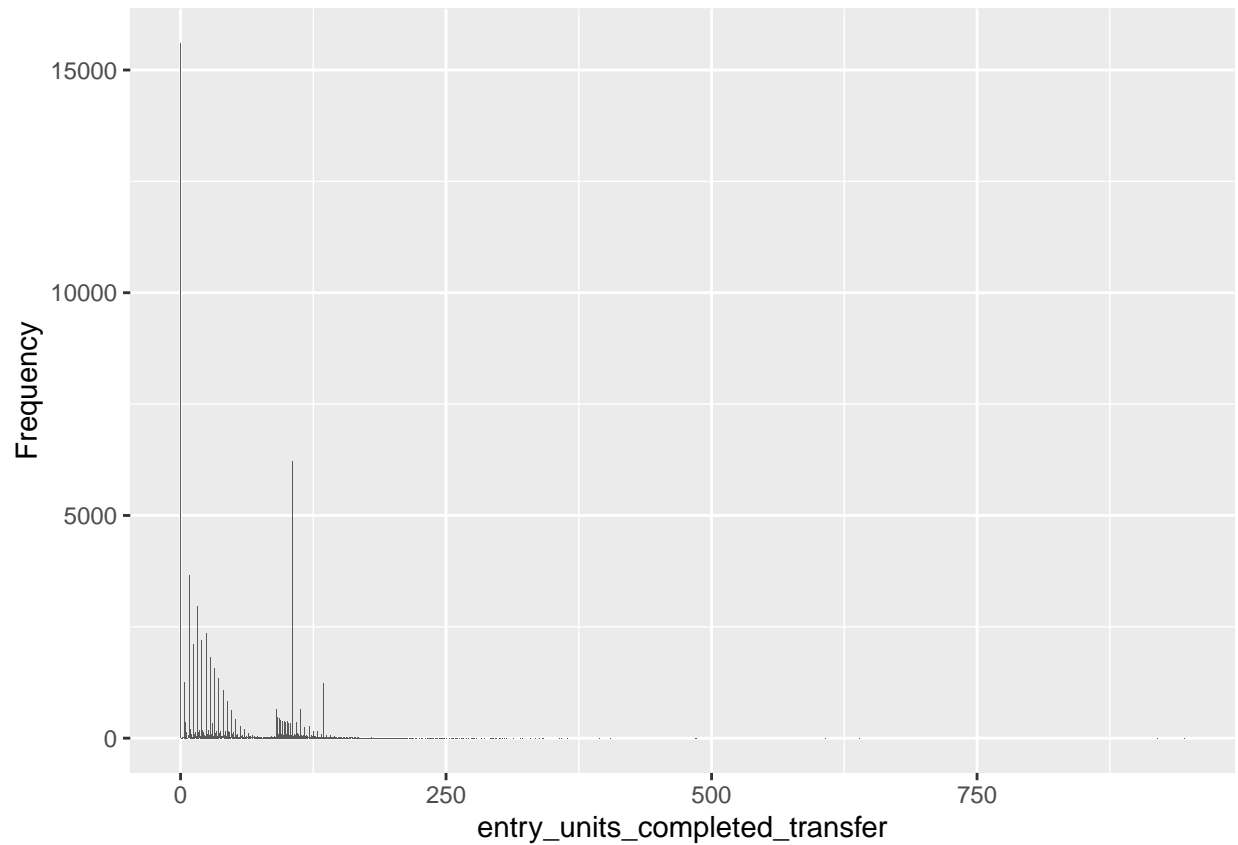


```
[1] -----
[1] Variable: cal_res_at_app, type: character, requested: yes
[1] Values (4 unique): California Resident, NA, Bona-Fide Resident, Non-California Resident
[1] Missing: 33.2%
Group.1 cal_res_at_app
1      2006      0.9989019608
2      2007      0.9985261606
3      2008      0.9955391478
4      2009      0.9974827991
5      2010      0.9970638232
6      2011      0.0000000000
7      2012      0.0002872738
8      2013      0.0006558237
9      2014      0.0036458333
10     2015      0.0031517902
11     2016      0.0047367854
12     2017      0.0014407739
13     2018      0.0102369450
```

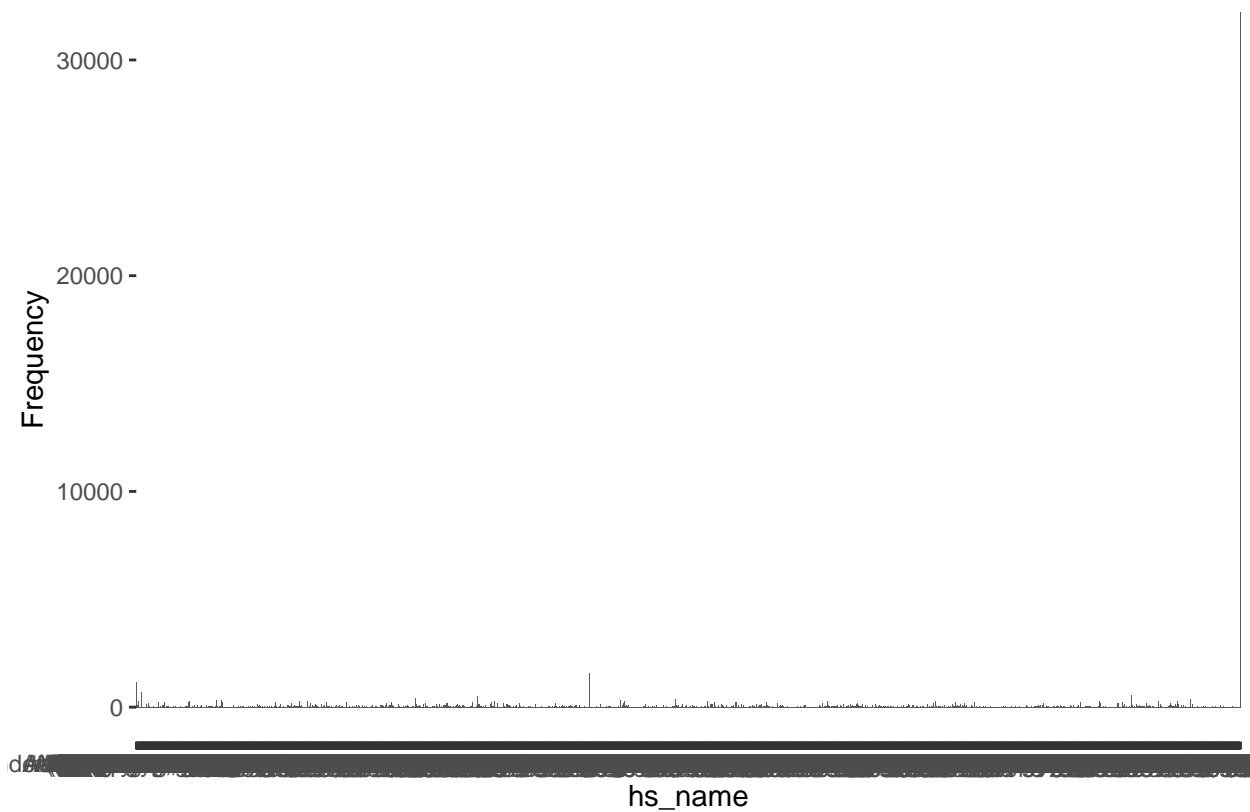


```
[1] -----
[1] Variable: entry_units_completed_transfer, type: numeric, requested: yes
[1] Values (825 unique): 209.2, NA, 0, 28, 32, ...
[1] Missing: 33.2%
  Group.1 entry_units_completed_transfer
1      2006                0.9989019608
2      2007                0.9985261606
3      2008                0.9955391478
4      2009                0.9974827991
5      2010                0.9970638232
6      2011                0.0000000000
7      2012                0.0002872738
8      2013                0.0001311647
9      2014                0.0036458333
10     2015                0.0031517902
11     2016                0.0047367854
12     2017                0.0013378615
13     2018                0.0095468139
```

Warning: Removed 32207 rows containing non-finite values ('stat_count()').



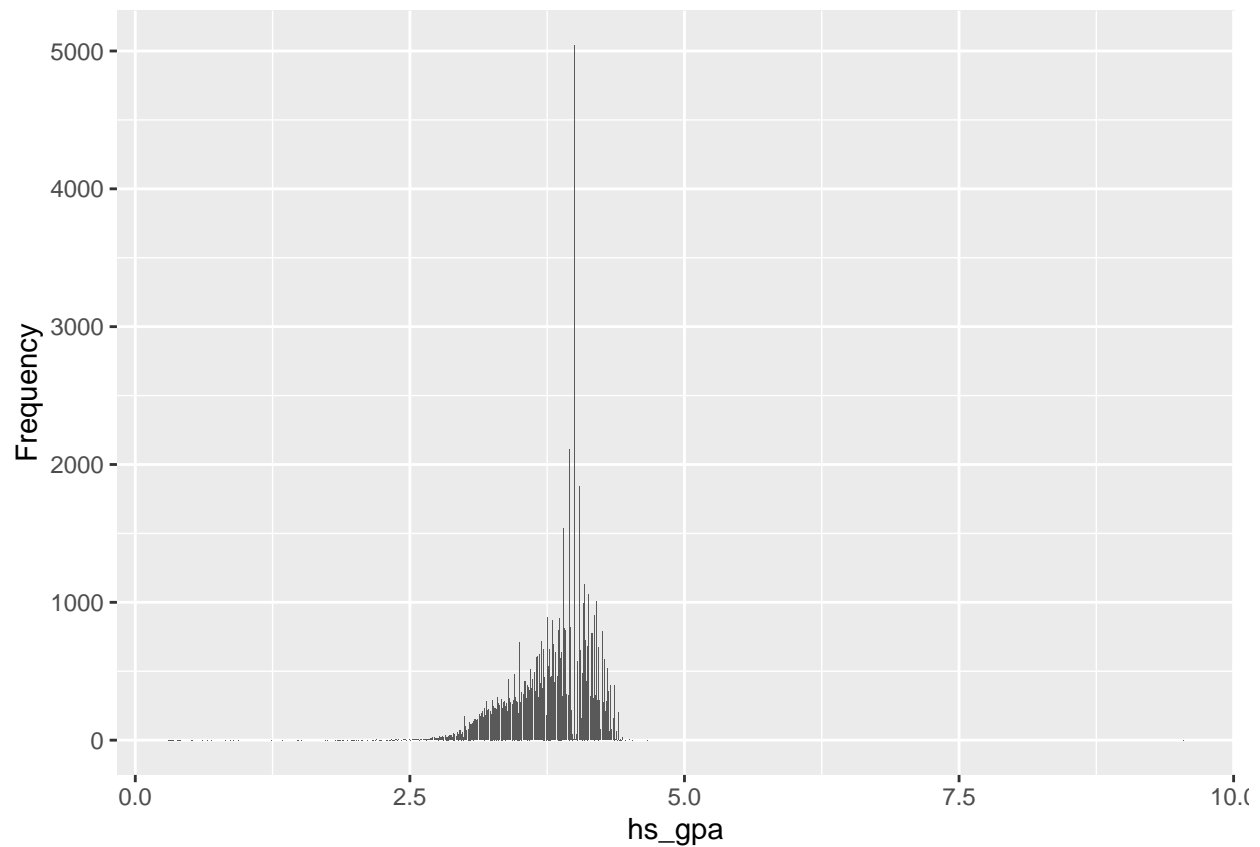
```
[1] -----
[1] Variable: hs_name, type: character, requested: yes
[1] Values (6426 unique): W H TAFT HS, NA, CROSSROADS CHRI, PALM DESERT HS, LELAND HS, ...
[1] Missing: 33.2%
  Group.1      hs_name
1      2006 0.9989019608
2      2007 0.9985261606
3      2008 0.9955391478
4      2009 0.9974827991
5      2010 0.9970638232
6      2011 0.0000000000
7      2012 0.0002872738
8      2013 0.0001311647
9      2014 0.0036458333
10     2015 0.0031517902
11     2016 0.0047367854
12     2017 0.0013378615
13     2018 0.0095468139
```



```
[1] -----
[1] Variable: hs_gpa, type: numeric, requested: yes
[1] Values (268 unique): 3.27, NA, 3.78, 4.1599998, 4.0799999, ...
[1] Missing: 34.5%
```

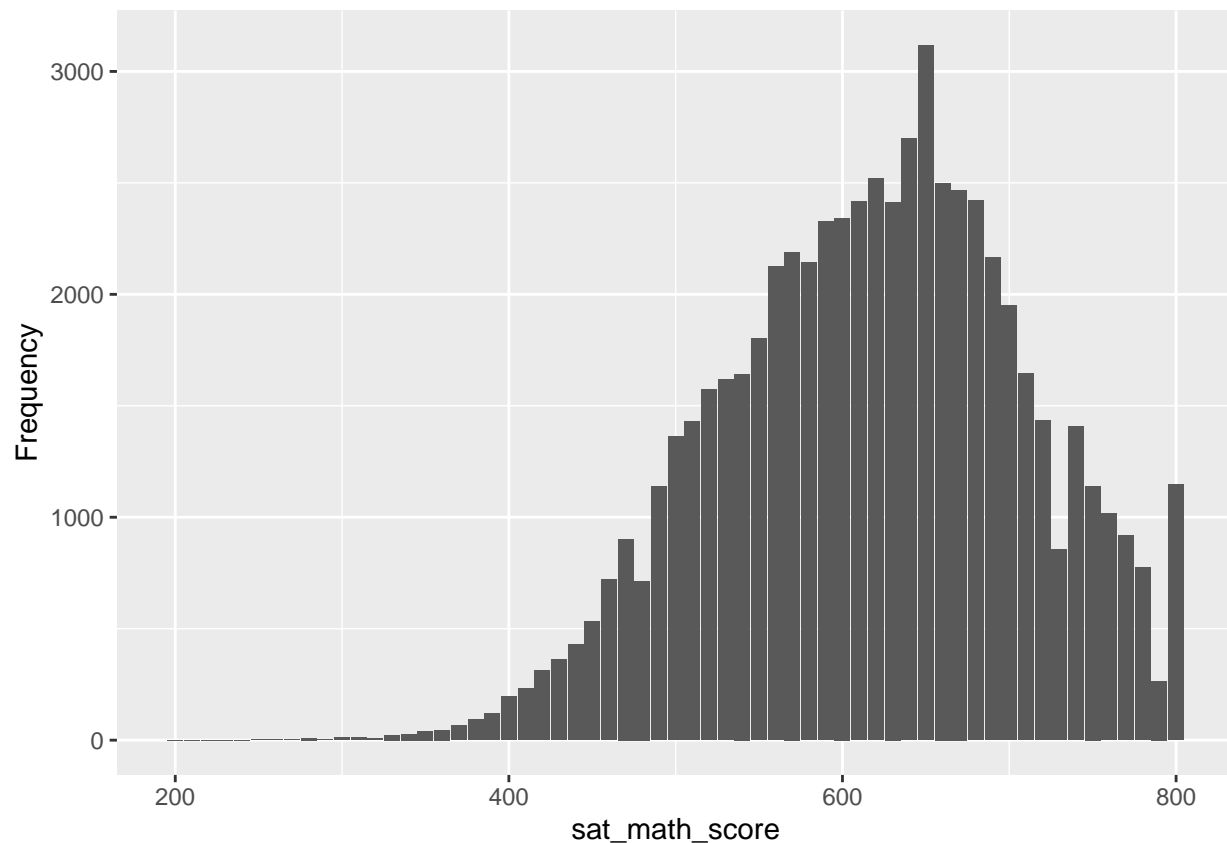
	Group.1	hs_gpa
1	2006	0.99905882
2	2007	0.99852616
3	2008	0.99553915
4	2009	0.99765061
5	2010	0.99737289
6	2011	0.02818697
7	2012	0.02212008
8	2013	0.02033054
9	2014	0.02565104
10	2015	0.02092789
11	2016	0.01980838
12	2017	0.01440774
13	2018	0.03485162

Warning: Removed 33513 rows containing non-finite values ('stat_count()').



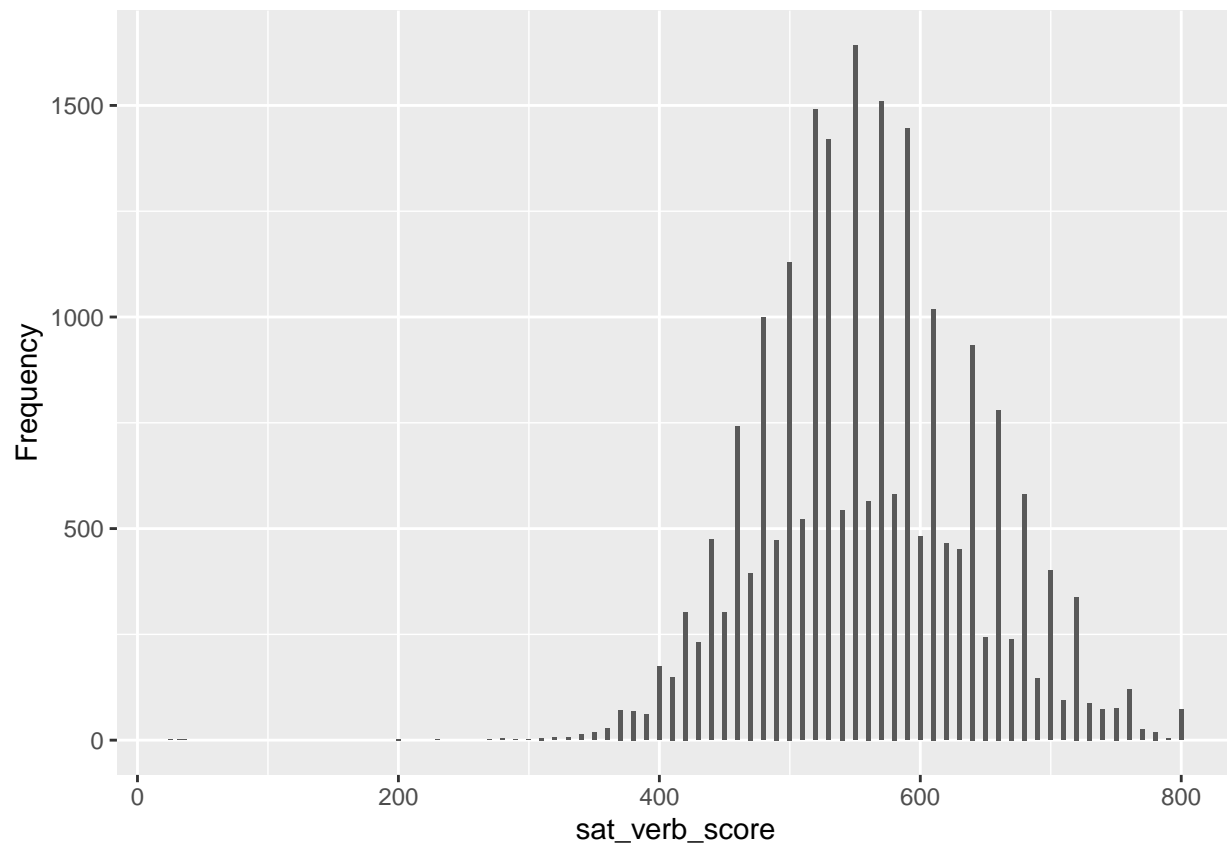
```
[1] -----  
[1] Variable: sat_math_score, type: numeric, requested: yes  
[1] Values (62 unique): NA, 630, 620, 780, 650, ...  
[1] Missing: 36.3%  
[1] Most missing: 2010 43.1%, Least missing: 2006 19.9%
```

Warning: Removed 35209 rows containing non-finite values ('stat_count()').



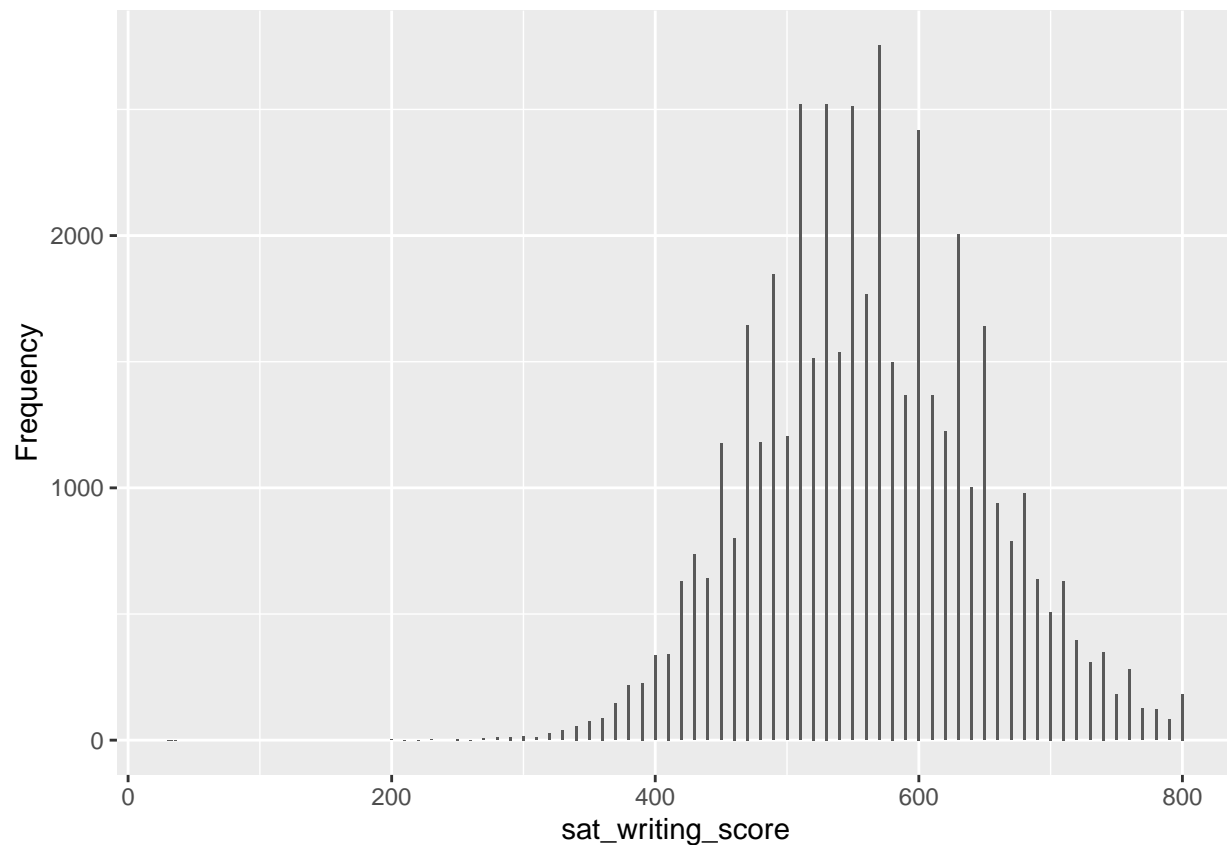
```
[1] is used in feature engineering
[1] -----
[1] Variable: sat_verb_score, type: numeric, requested: yes
[1] Values (60 unique): NA, 500, 720, 680, 590, ...
[1] Missing: 77.3%
  Group.1 sat_verb_score
1    2006      1.0000000
2    2007      0.9998526
3    2008      1.0000000
4    2009      1.0000000
5    2010      1.0000000
6    2011      0.9915014
7    2012      0.9900891
8    2013      0.9786201
9    2014      0.8975260
10   2015      0.4750378
11   2016      0.4215739
12   2017      0.4075332
13   2018      0.3489763
```

Warning: Removed 75015 rows containing non-finite values ('stat_count()').



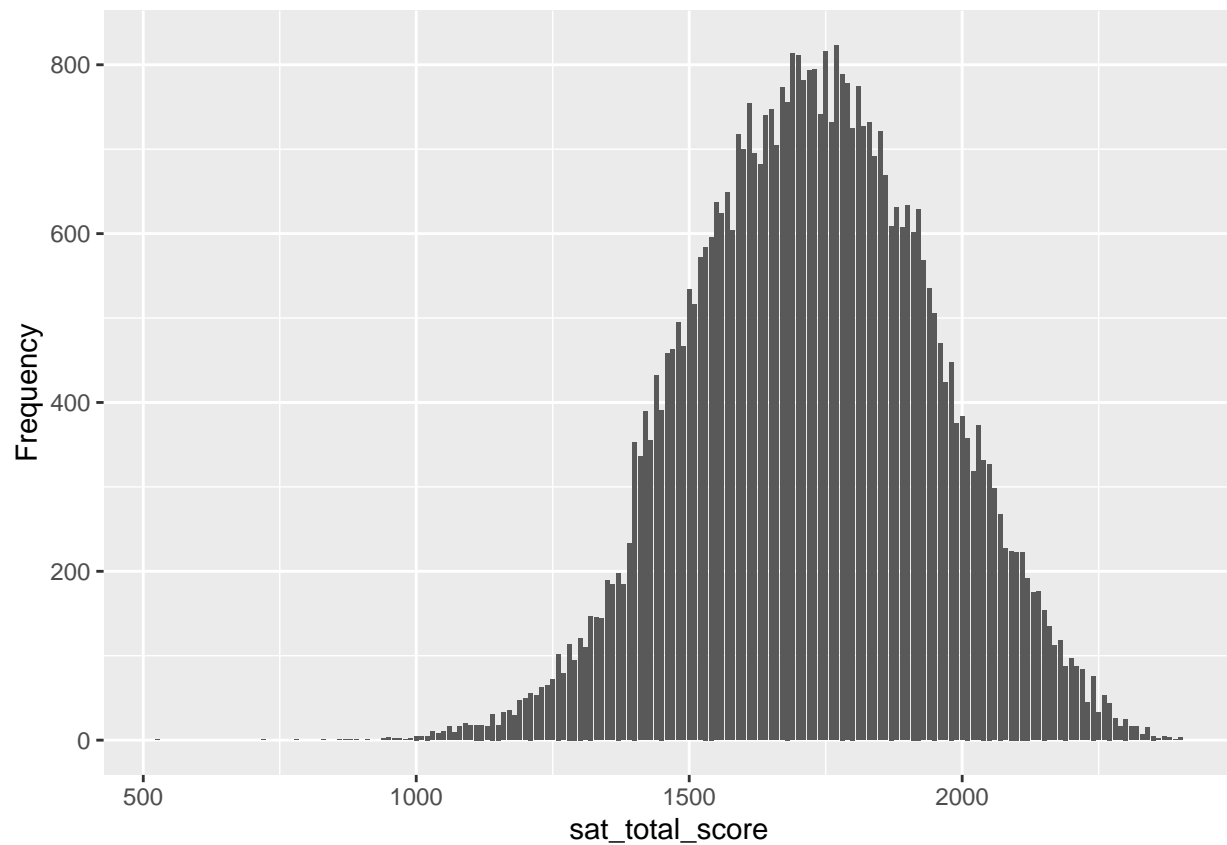
```
[1] is used in feature engineering
[1] -----
[1] Variable: sat_writing_score, type: numeric, requested: yes
[1] Values (64 unique): NA, 510, 680, 740, 600, ...
[1] Missing: 53%
  Group.1 sat_writing_score
1    2006      0.9995294
2    2007      0.9994105
3    2008      0.9973850
4    2009      0.9984897
5    2010      0.9984546
6    2011      0.2338527
7    2012      0.2318299
8    2013      0.2692812
9    2014      0.2865885
10   2015      0.2895865
11   2016      0.3257617
12   2017      0.3613255
13   2018      0.3484012
```

Warning: Removed 51430 rows containing non-finite values ('stat_count()').



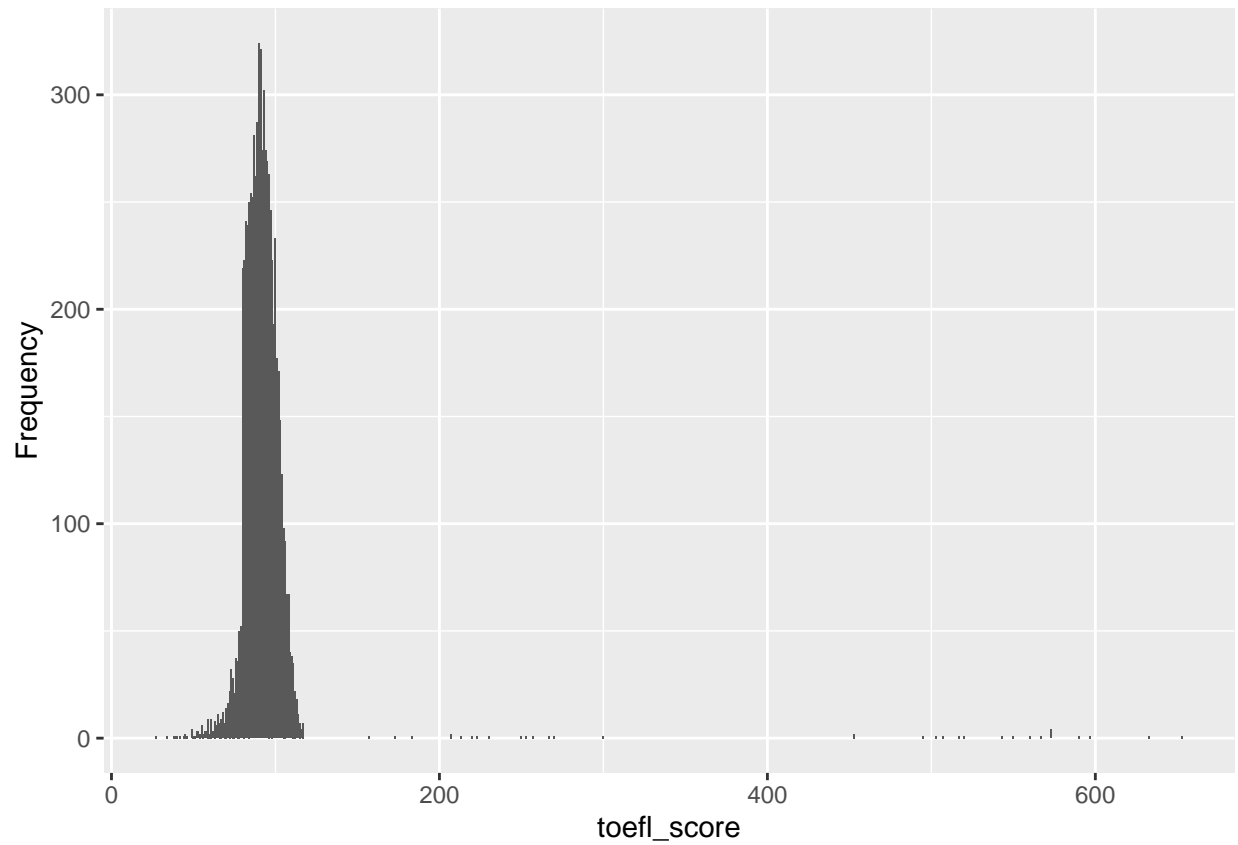
```
[1] is used in feature engineering
[1] -----
[1] Variable: sat_total_score, type: numeric, requested: yes
[1] Values (157 unique): NA, 1640, 2020, 2200, 1930, ...
[1] Missing: 53%
  Group.1 sat_total_score
1    2006      0.9993725
2    2007      0.9994105
3    2008      0.9972312
4    2009      0.9984897
5    2010      0.9984546
6    2011      0.2327195
7    2012      0.2311117
8    2013      0.2682319
9    2014      0.2865885
10   2015      0.2894604
11   2016      0.3256540
12   2017      0.3614284
13   2018      0.3511617
```

Warning: Removed 51430 rows containing non-finite values ('stat_count()').



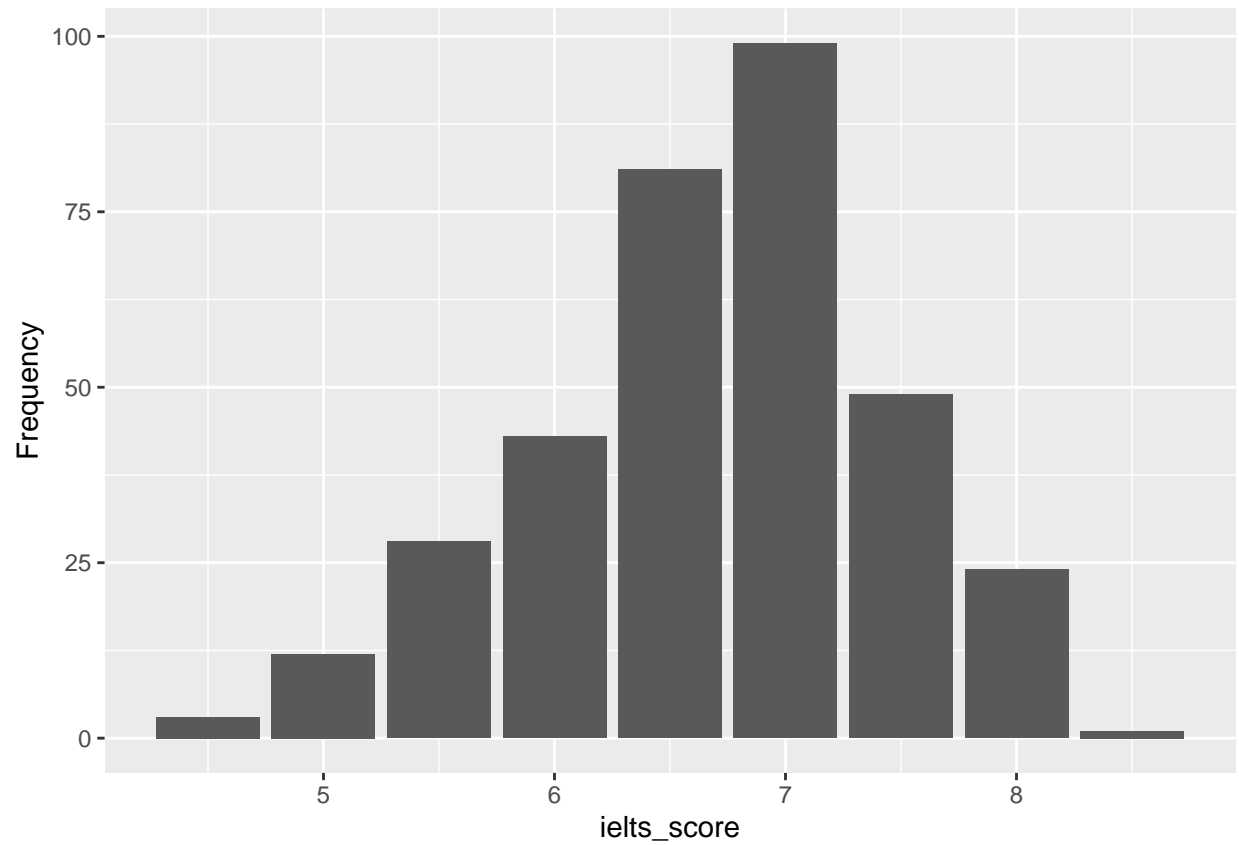
```
[1] is used in feature engineering
[1] -----
[1] Variable: toefl_score, type: numeric, requested: yes
[1] Values (108 unique): NA, 95, 104, 93, 90, ...
[1] Missing: 92.8%
[1] Most missing: 2006 100%, Least missing: 2016 84.3%
```

Warning: Removed 90031 rows containing non-finite values ('stat_count()').

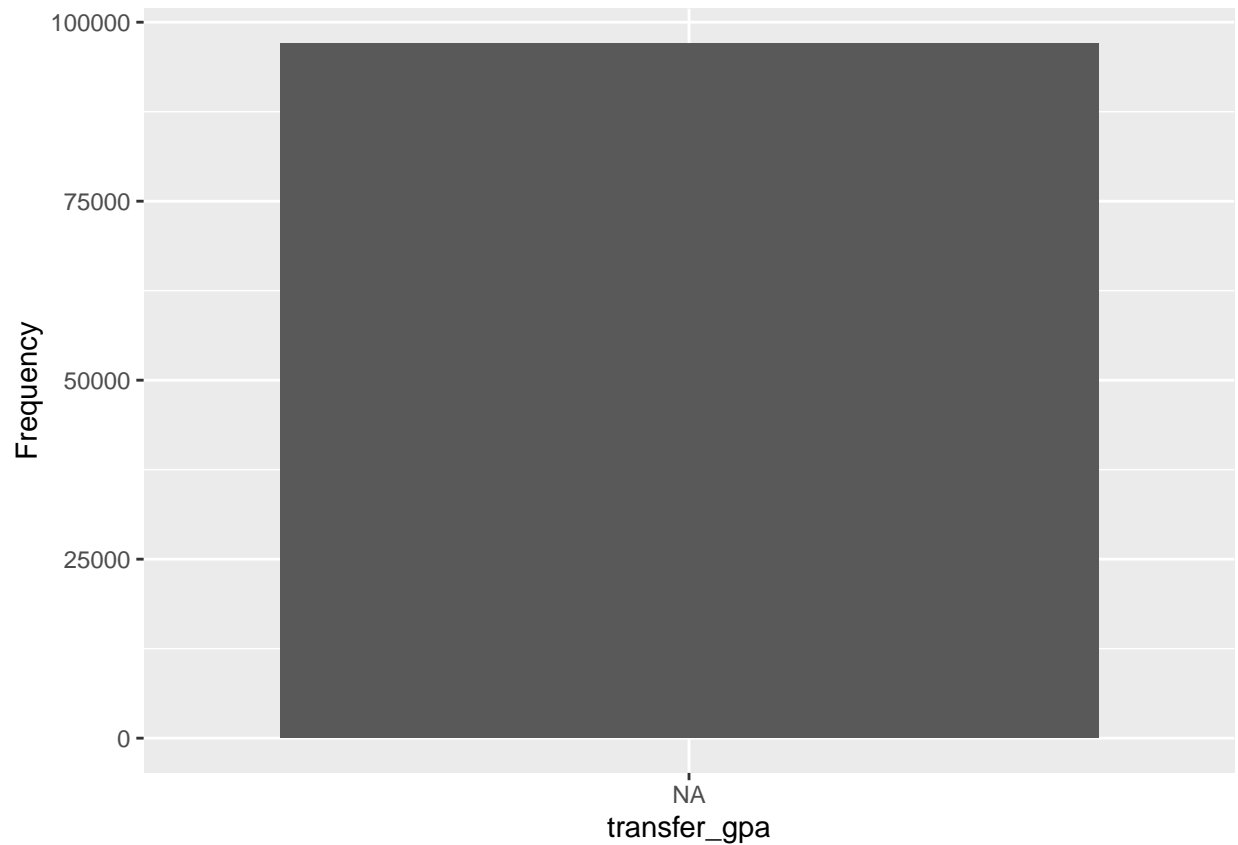


```
[1] -----  
[1] Variable: ielts_score, type: numeric, requested: yes  
[1] Values (10 unique): NA, 7, 8, 7.5, 6, ...  
[1] Missing: 99.6%
```

Warning: Removed 96709 rows containing non-finite values ('stat_count()').

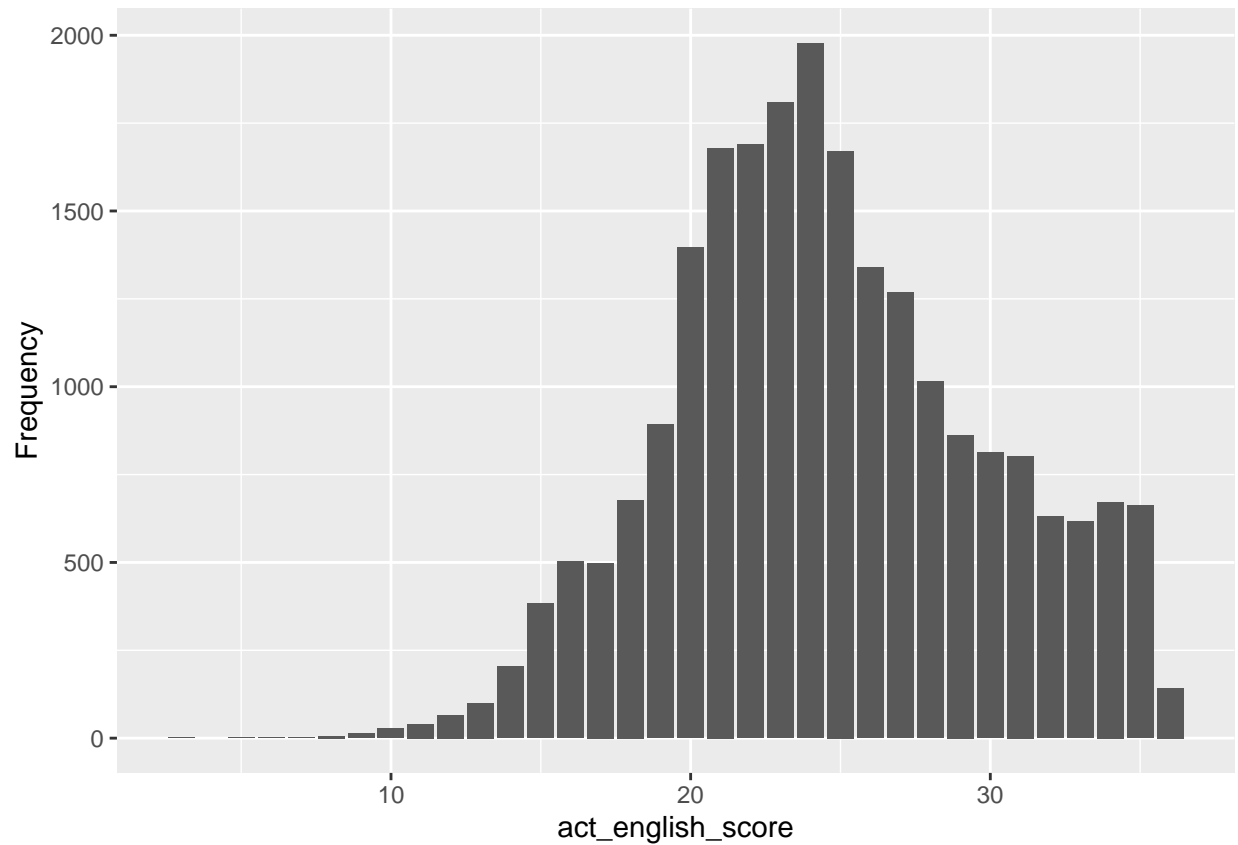


```
[1] -----  
[1] Variable: transfer_gpa, type: logical, requested: yes  
[1] Values (1 unique): NA  
[1] Missing: 100%
```



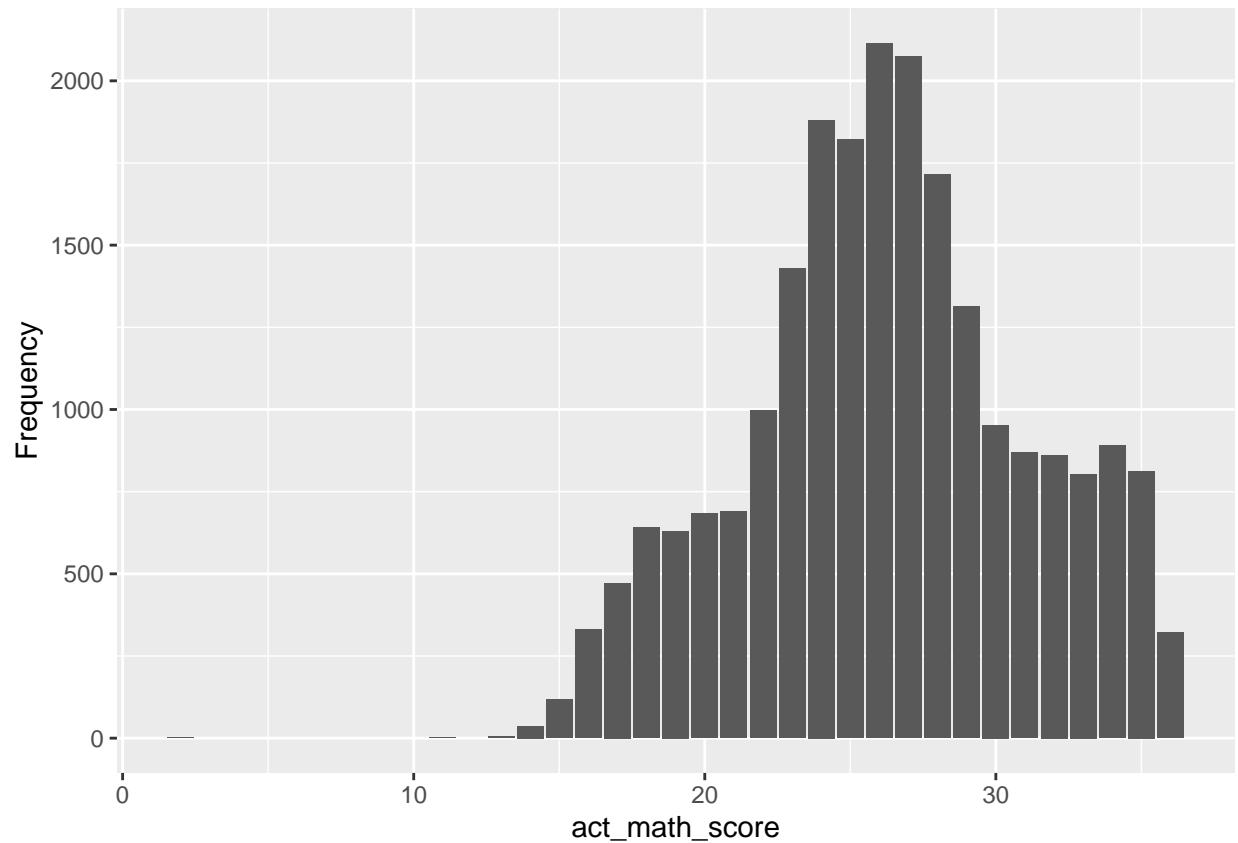
```
[1] -----  
[1] Variable: act_english_score, type: numeric, requested: yes  
[1] Values (34 unique): NA, 33, 26, 27, 23, ...  
[1] Missing: 76.9%  
[1] Most missing: 2006 100%, Least missing: 2016 62.5%
```

Warning: Removed 74586 rows containing non-finite values ('stat_count()').



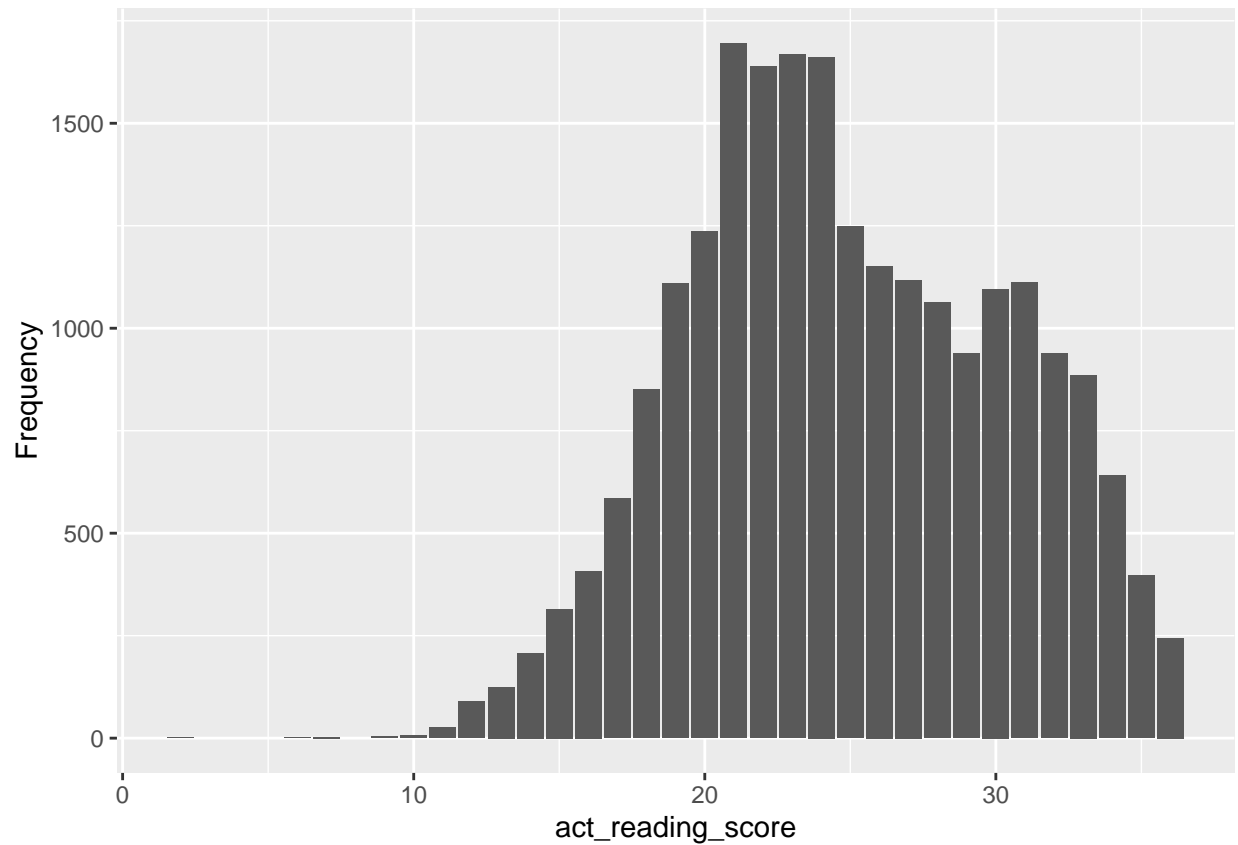
```
[1] is used in feature engineering
[1] -----
[1] Variable: act_math_score, type: numeric, requested: yes
[1] Values (27 unique): NA, 33, 26, 28, 14, ...
[1] Missing: 76.9%
[1] Most missing: 2006 100%, Least missing: 2016 62.5%
```

Warning: Removed 74586 rows containing non-finite values ('stat_count()').



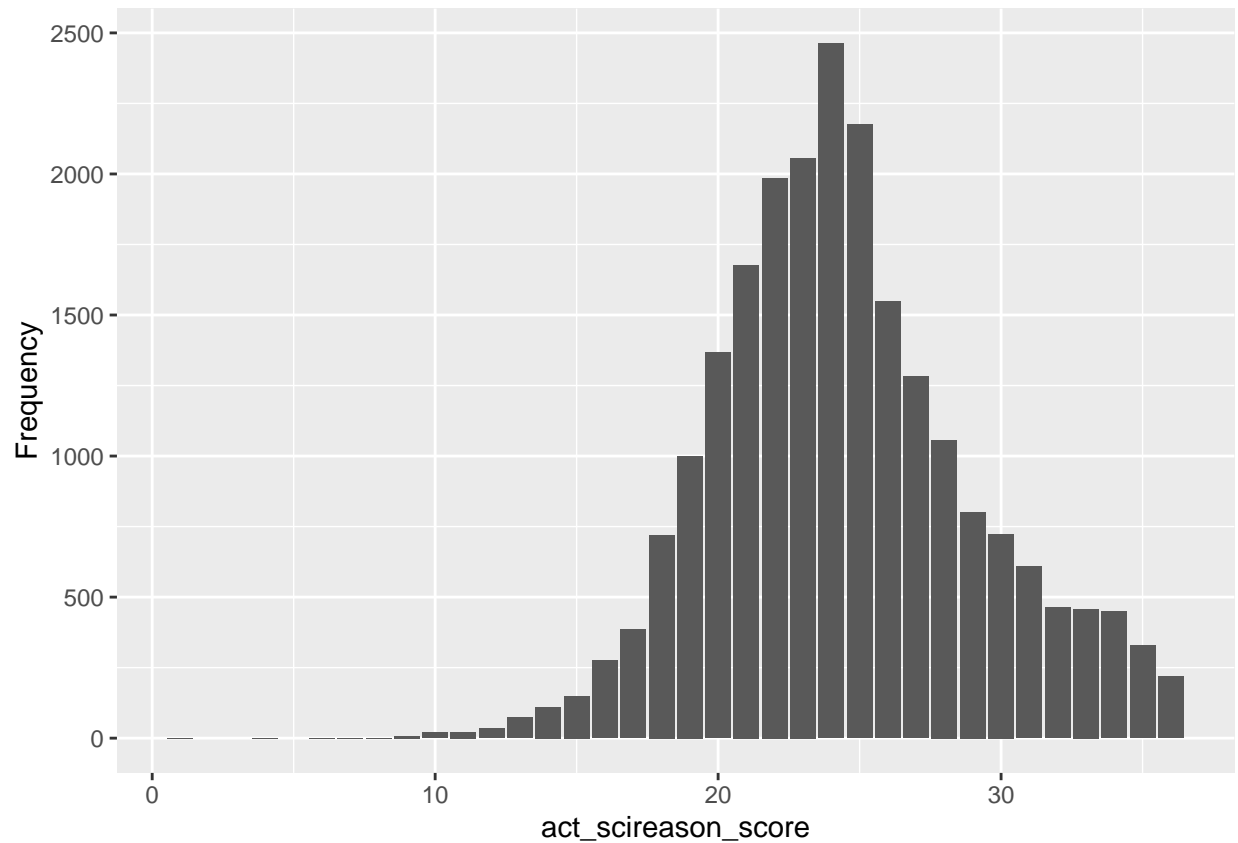
```
[1] is used in feature engineering
[1] -----
[1] Variable: act_reading_score, type: numeric, requested: yes
[1] Values (32 unique): NA, 35, 22, 34, 32, ...
[1] Missing: 76.9%
[1] Most missing: 2006 100%, Least missing: 2016 62.5%
```

Warning: Removed 74586 rows containing non-finite values ('stat_count()').



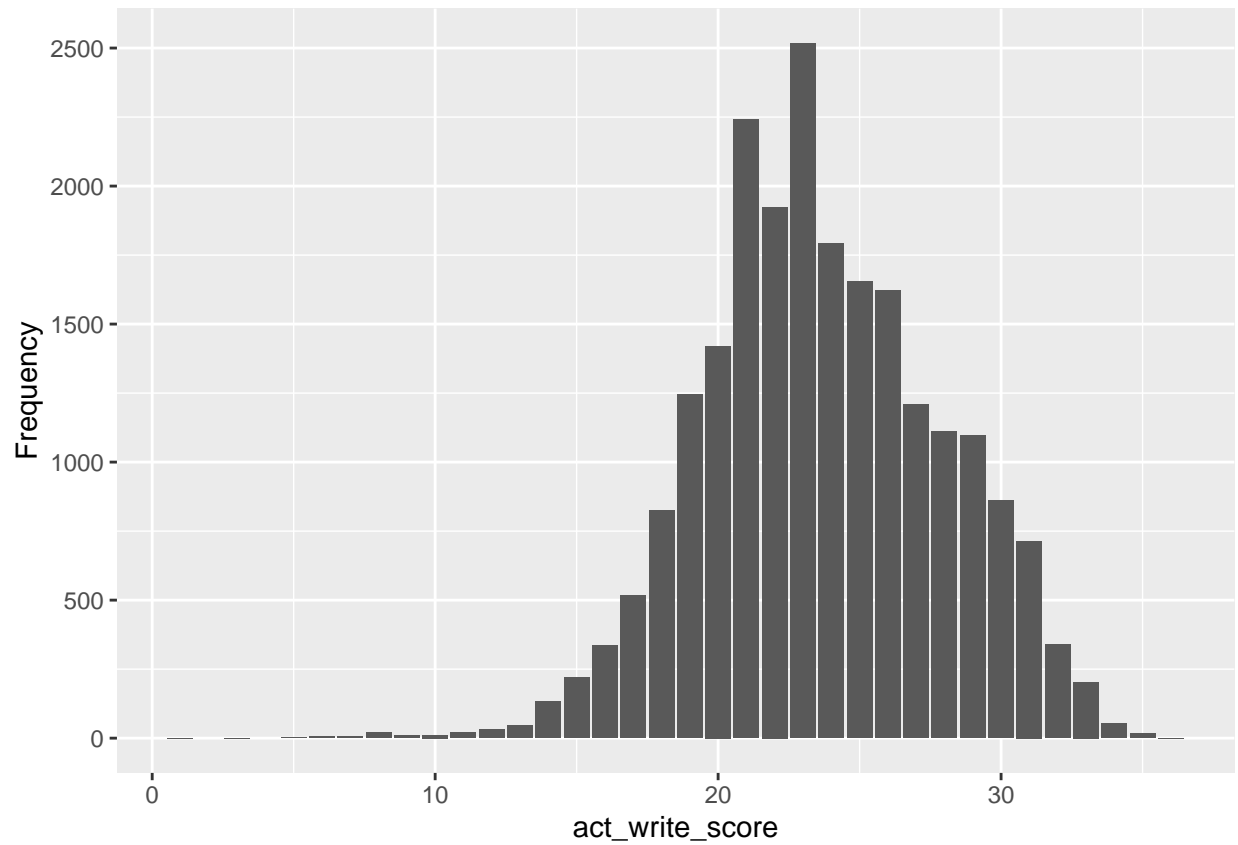
```
[1] is used in feature engineering
[1] -----
[1] Variable: act_scireason_score, type: numeric, requested: yes
[1] Values (34 unique): NA, 35, 25, 24, 30, ...
[1] Missing: 76.9%
[1] Most missing: 2006 100%, Least missing: 2016 62.5%
```

Warning: Removed 74586 rows containing non-finite values ('stat_count()').

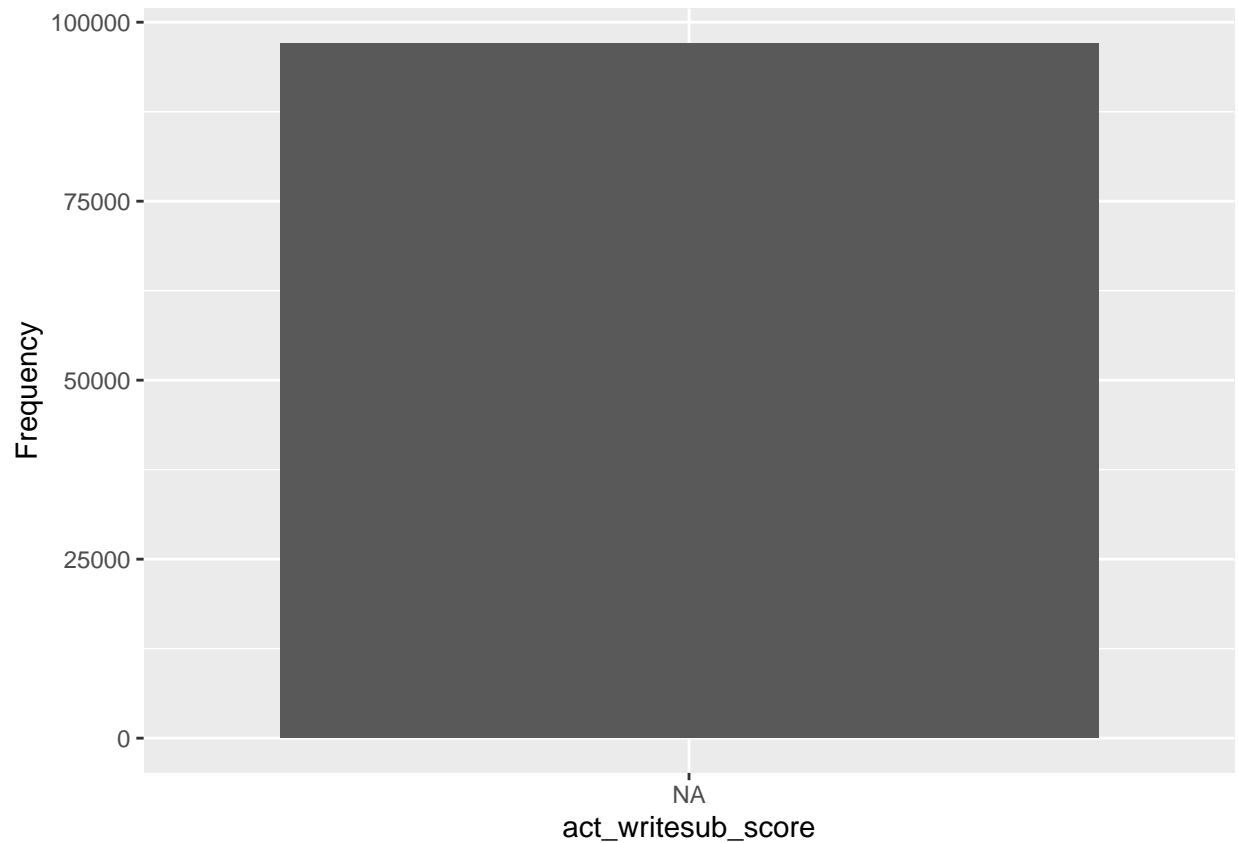


```
[1] is used in feature engineering
[1] -----
[1] Variable: act_write_score, type: numeric, requested: yes
[1] Values (35 unique): NA, 32, 24, 29, 21, ...
[1] Missing: 77.1%
[1] Most missing: 2007 100%, Least missing: 2016 63%
```

Warning: Removed 74842 rows containing non-finite values ('stat_count()').

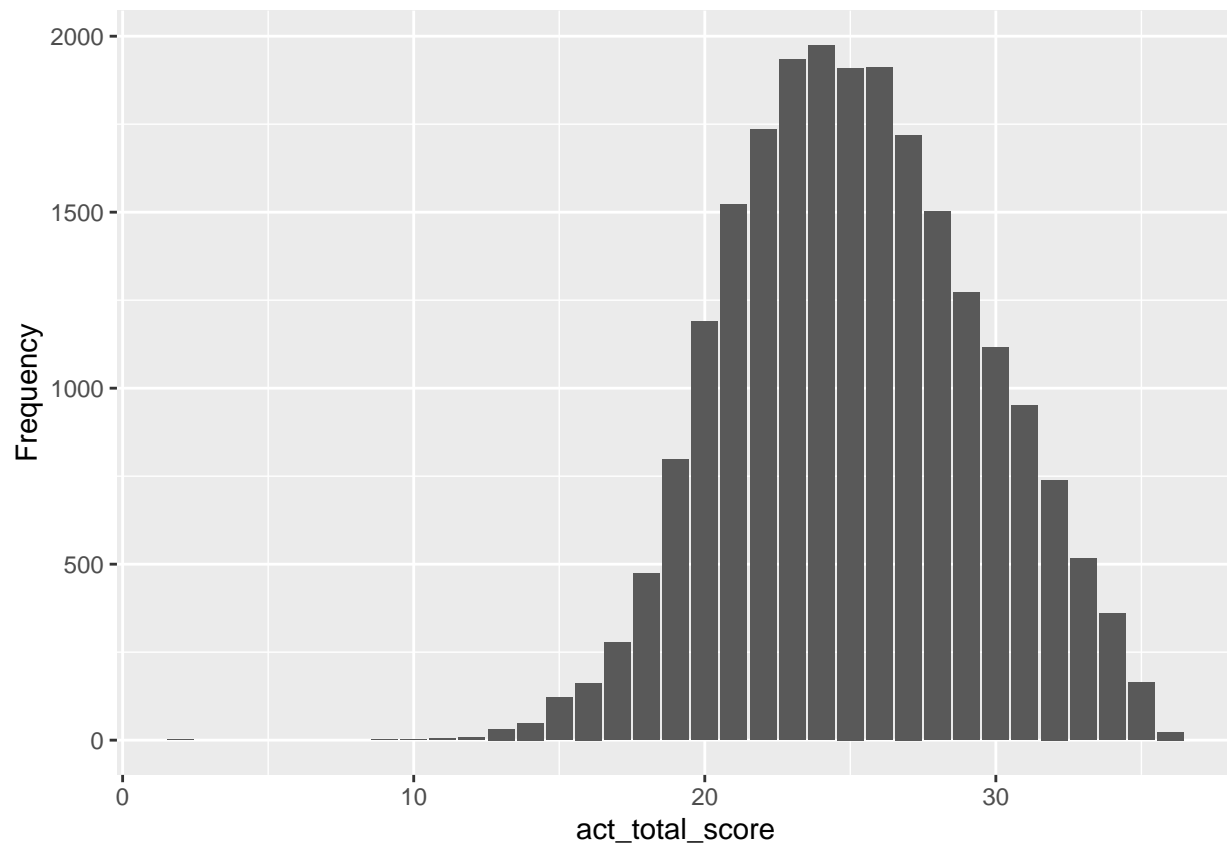


```
[1] is used in feature engineering
[1] -----
[1] Variable: act_writesub_score, type: logical, requested: yes
[1] Values (1 unique): NA
[1] Missing: 100%
```

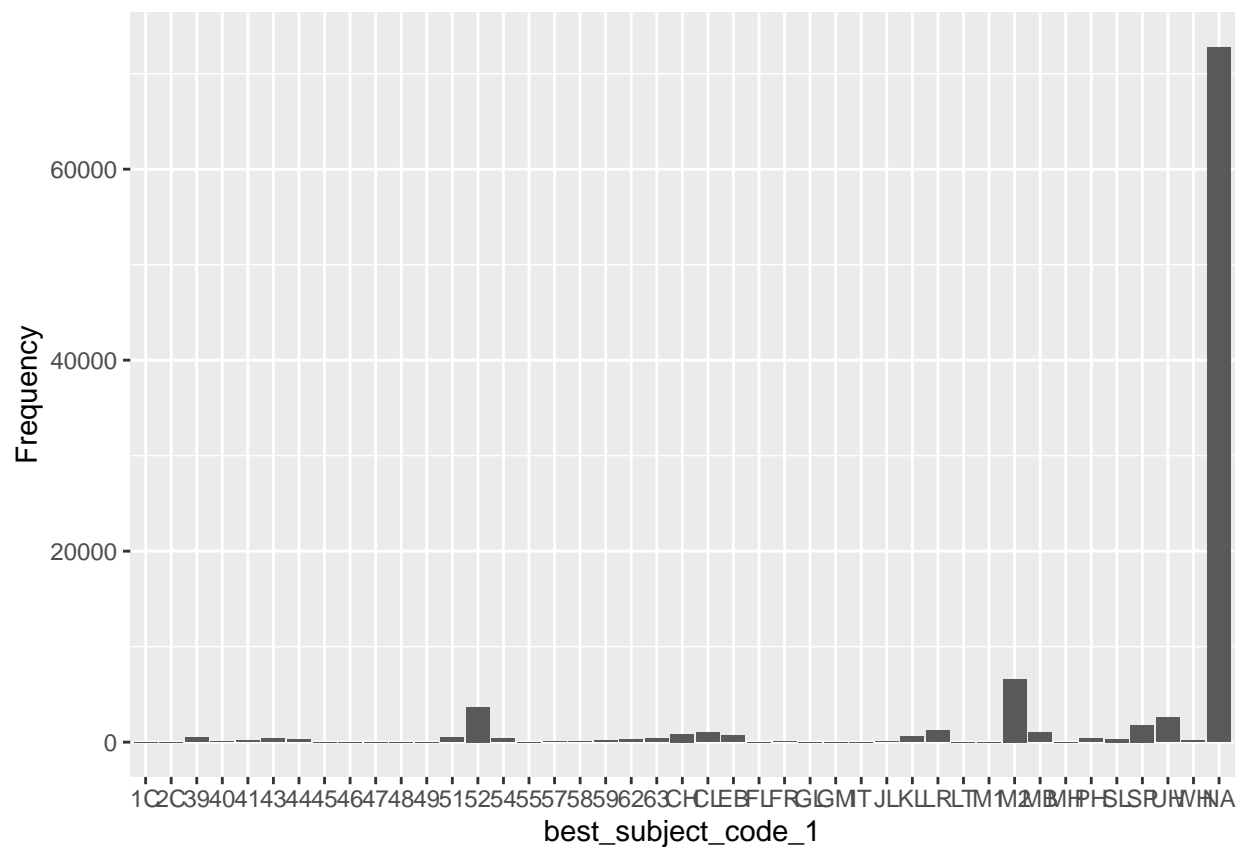


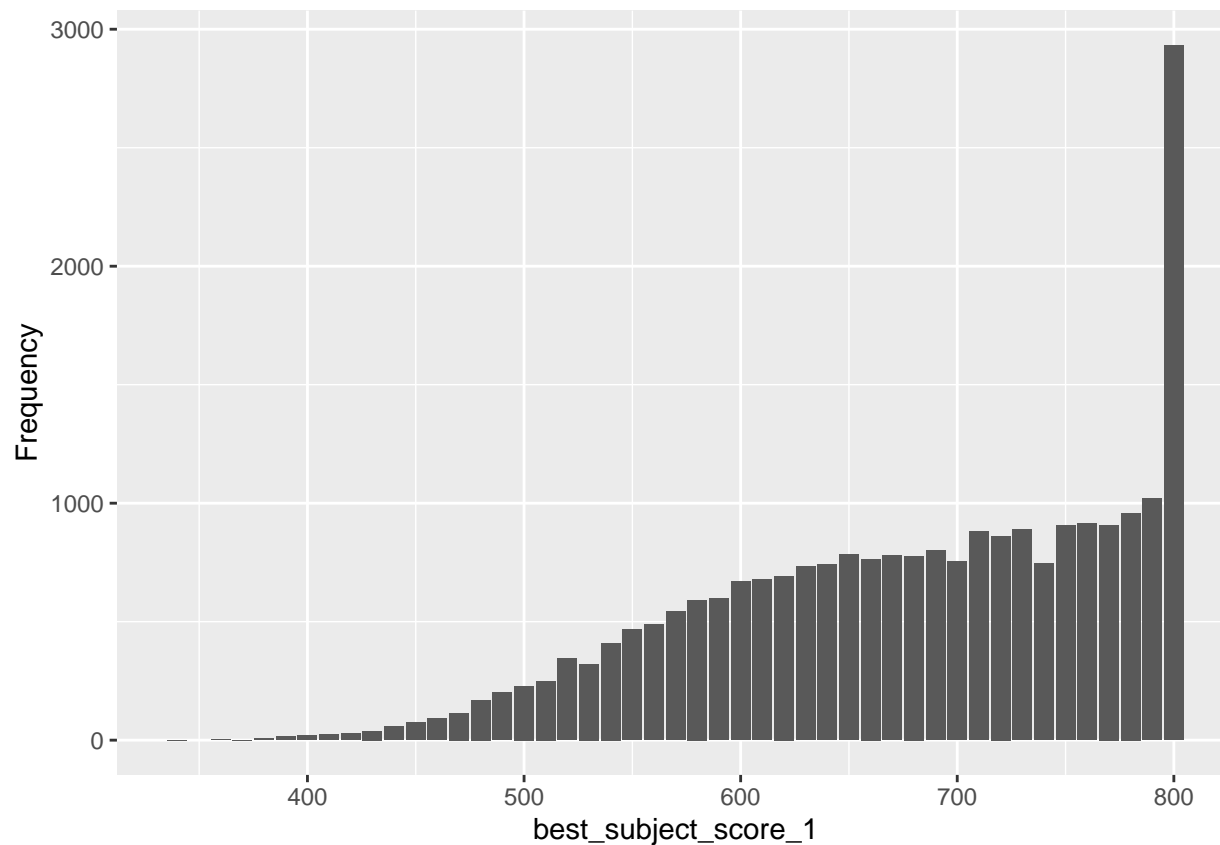
```
[1] is used in feature engineering
[1] -----
[1] Variable: act_total_score, type: numeric, requested: yes
[1] Values (30 unique): NA, 34, 25, 28, 31, ...
[1] Missing: 76.9%
[1] Most missing: 2006 100%, Least missing: 2016 62.5%
```

Warning: Removed 74586 rows containing non-finite values ('stat_count()').

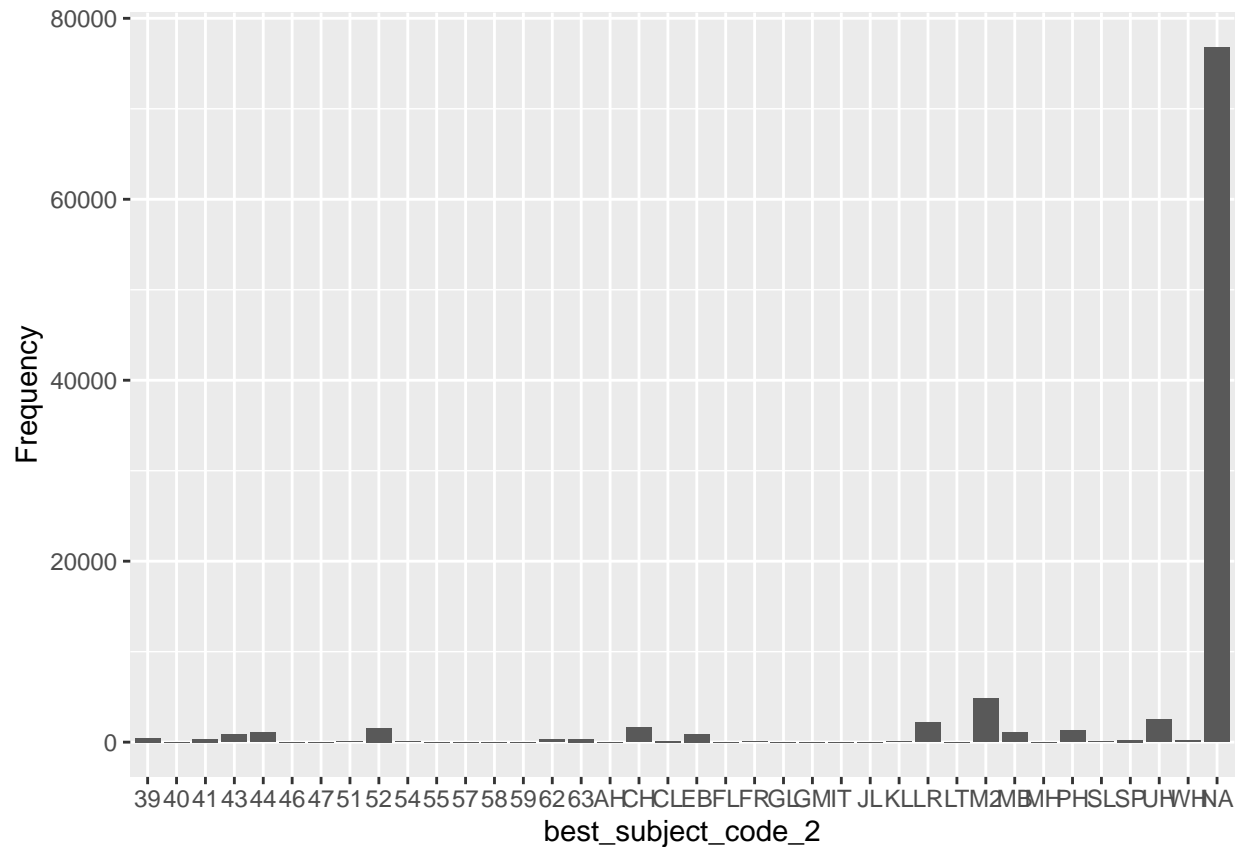


```
[1] is used in feature engineering
[1] -----
[1] Variable: best_subject_code_1, type: character, requested: yes
[1] Values (43 unique): NA, 54, 52, 51, M2, ...
[1] Missing: 75%
Group.1 best_subject_code_1
1      2006      0.9995294
2      2007      0.9994105
3      2008      0.9973850
4      2009      0.9984897
5      2010      0.9984546
6      2011      0.2201133
7      2012      0.5435220
8      2013      0.6130640
9      2014      0.6559896
10     2015      0.7010842
11     2016      0.7072882
12     2017      0.7535248
13     2018      0.7135956
```



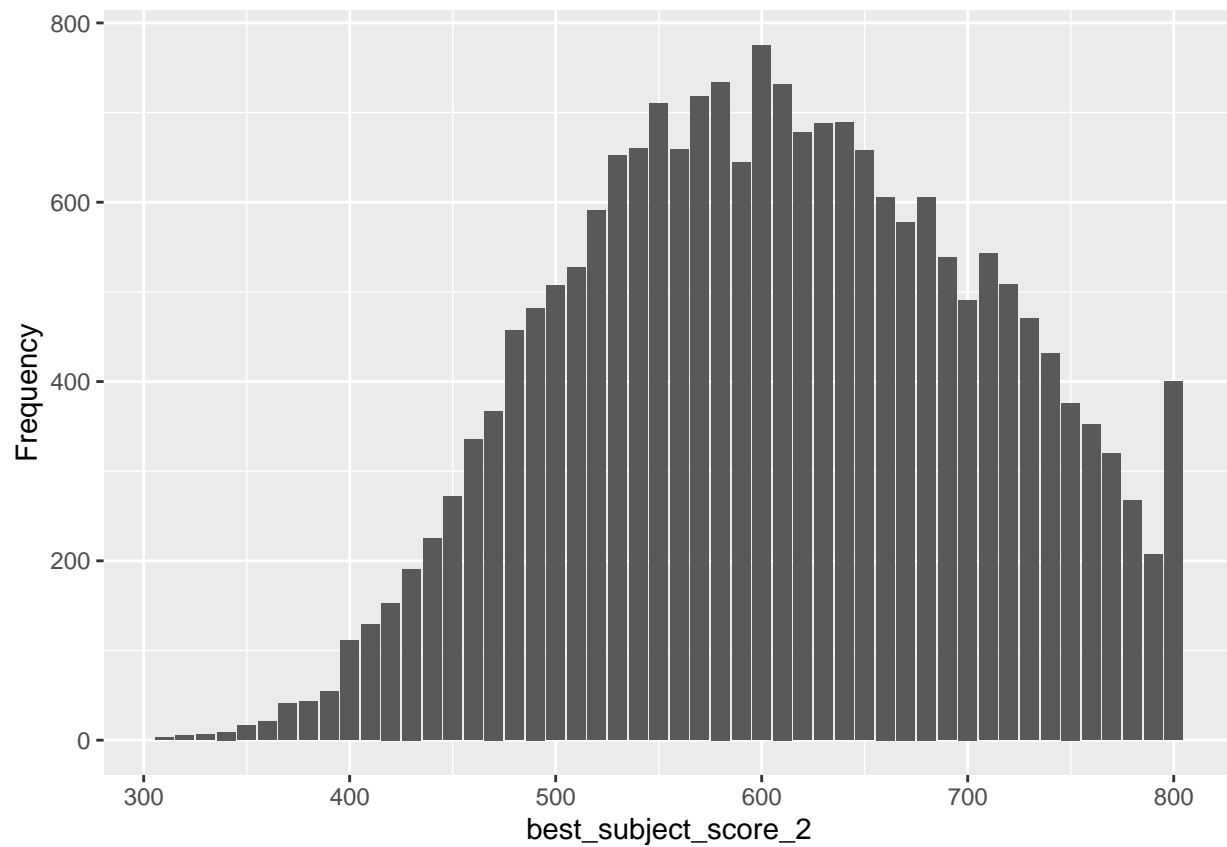


```
[1] -----
[1] Variable: best_subject_code_2, type: character, requested: yes
[1] Values (38 unique): NA, 39, 44, 41, SL, ...
[1] Missing: 79.1%
Group.1 best_subject_code_2
1      2006      0.9995294
2      2007      0.9994105
3      2008      0.9973850
4      2009      0.9984897
5      2010      0.9984546
6      2011      0.2223796
7      2012      0.6128986
8      2013      0.6785152
9      2014      0.7248698
10     2015      0.7656329
11     2016      0.7839380
12     2017      0.8227848
13     2018      0.7841040
```

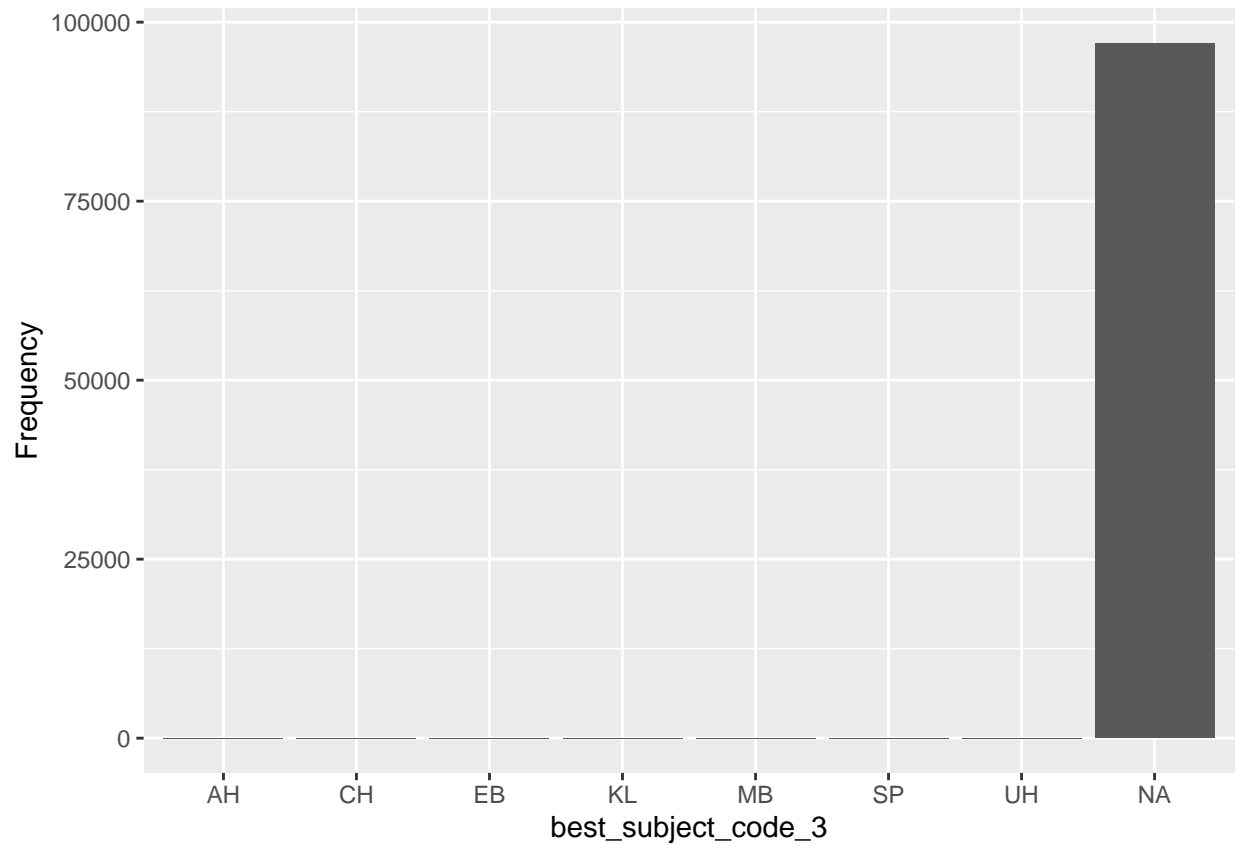


```
[1] -----
[1] Variable: best_subject_score_2, type: integer, requested: yes
[1] Values (51 unique): NA, 680, 670, 630, 710, ...
[1] Missing: 79.1%
Group.1 best_subject_score_2
1      2006      0.9995294
2      2007      0.9994105
3      2008      0.9973850
4      2009      0.9984897
5      2010      0.9984546
6      2011      0.2223796
7      2012      0.6128986
8      2013      0.6785152
9      2014      0.7248698
10     2015      0.7682804
11     2016      0.7840456
12     2017      0.8227848
13     2018      0.7841040
```

Warning: Removed 76814 rows containing non-finite values ('stat_count()').

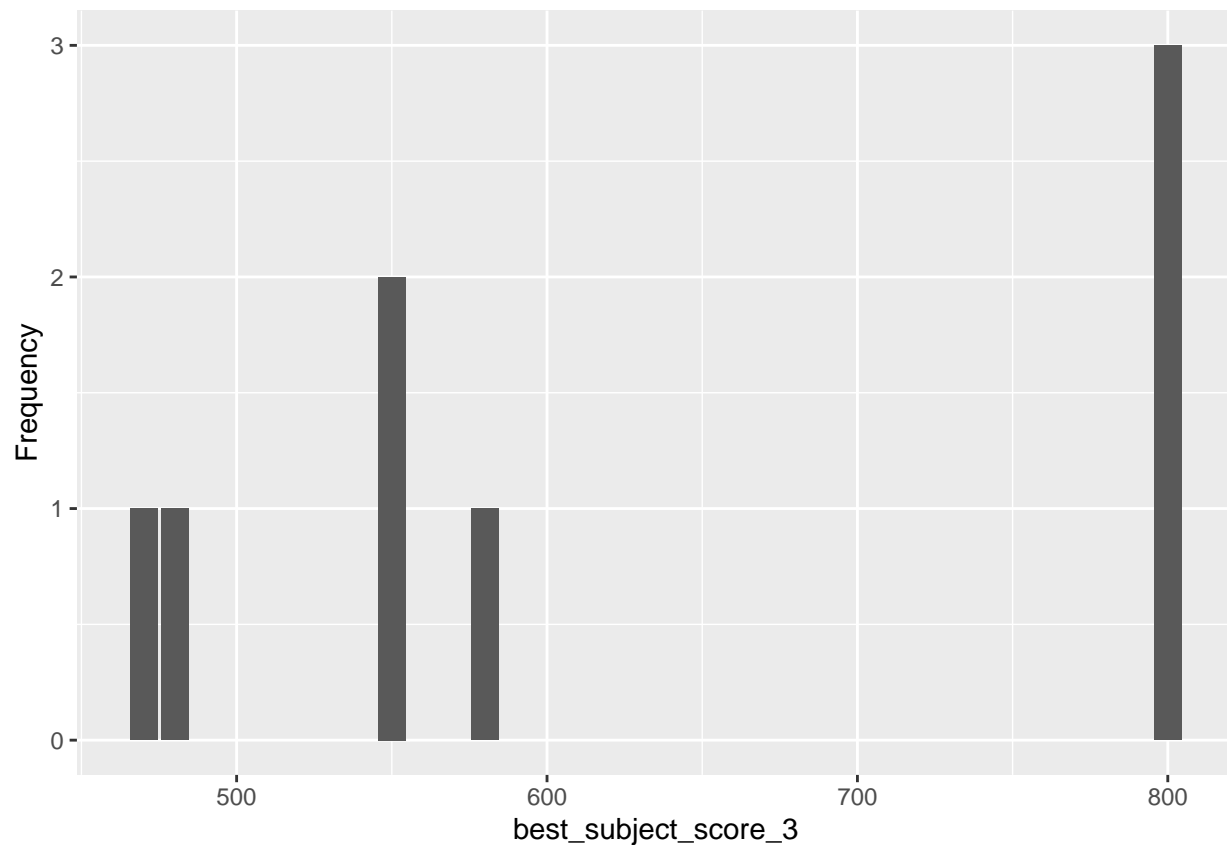


```
[1] -----  
[1] Variable: best_subject_code_3, type: character, requested: yes  
[1] Values (8 unique): NA, UH, MB, EB, KL, ...  
[1] Missing: 100%
```



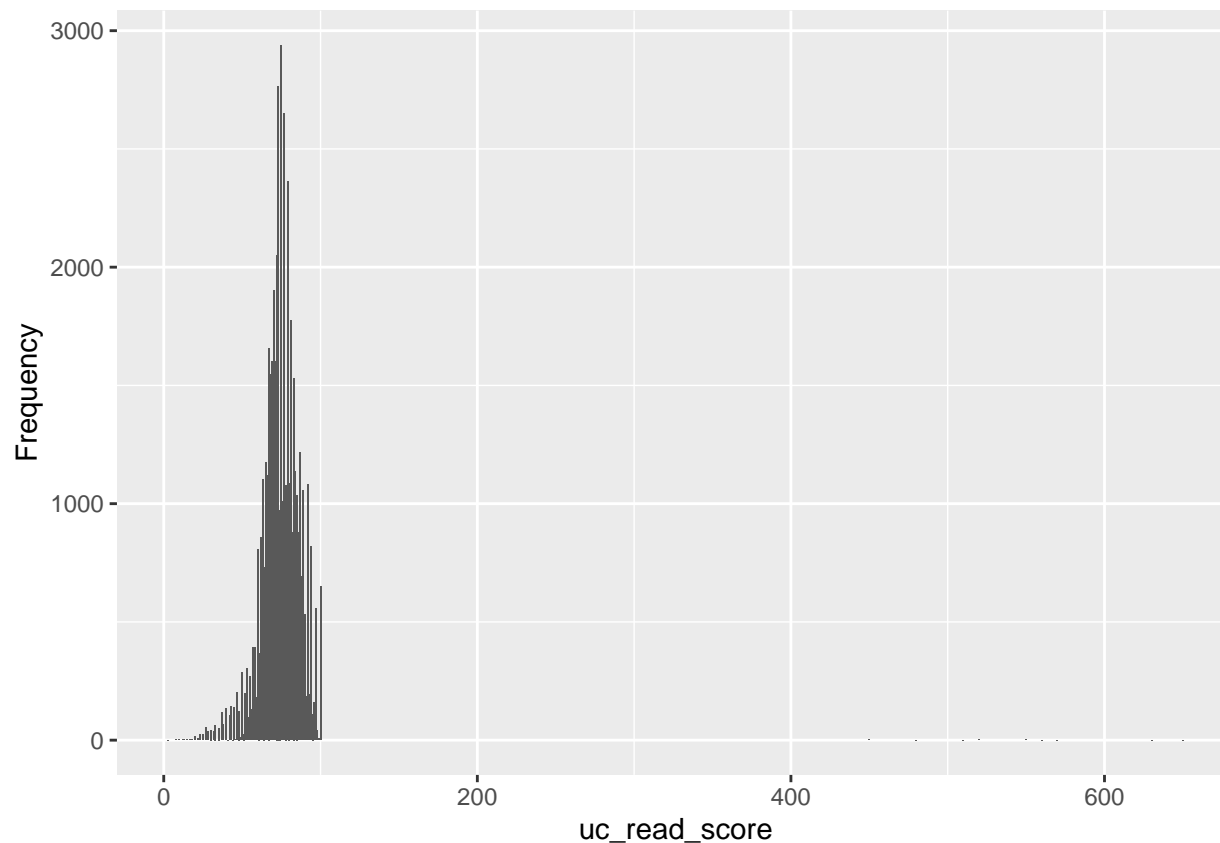
```
[1] -----  
[1] Variable: best_subject_score_3, type: integer, requested: yes  
[1] Values (6 unique): NA, 580, 470, 550, 800, ...  
[1] Missing: 100%
```

Warning: Removed 97041 rows containing non-finite values ('stat_count()').



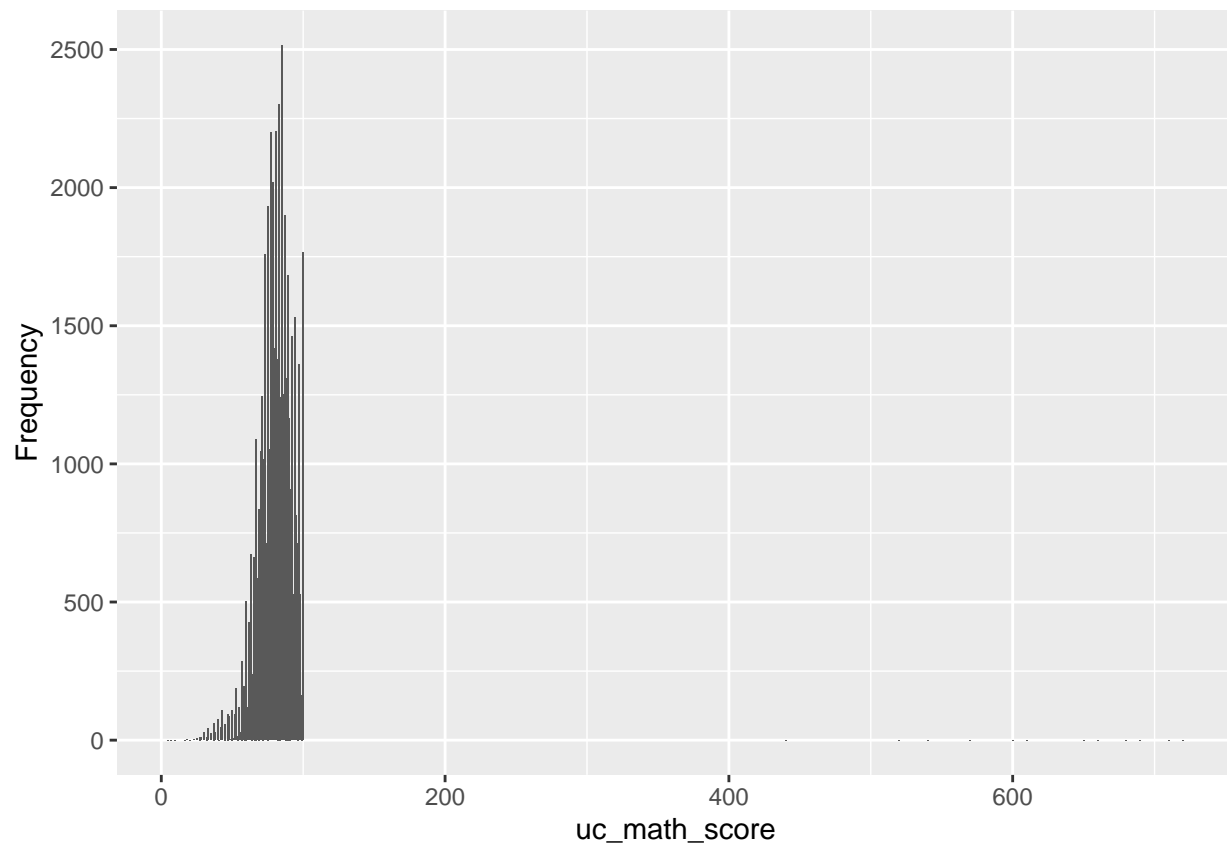
```
[1] -----
[1] Variable: uc_read_score, type: numeric, requested: yes
[1] Values (91 unique): NA, 70, 92, 88, 100, ...
[1] Missing: 48.9%
Group.1 uc_read_score
1      2006      0.9993725
2      2007      0.9994105
3      2008      0.9972312
4      2009      0.9984897
5      2010      0.9983001
6      2011      0.2114731
7      2012      0.2032462
8      2013      0.2370147
9      2014      0.2458333
10     2015      0.2257943
11     2016      0.2360857
12     2017      0.2615005
13     2018      0.2597193
```

Warning: Removed 47431 rows containing non-finite values ('stat_count()').



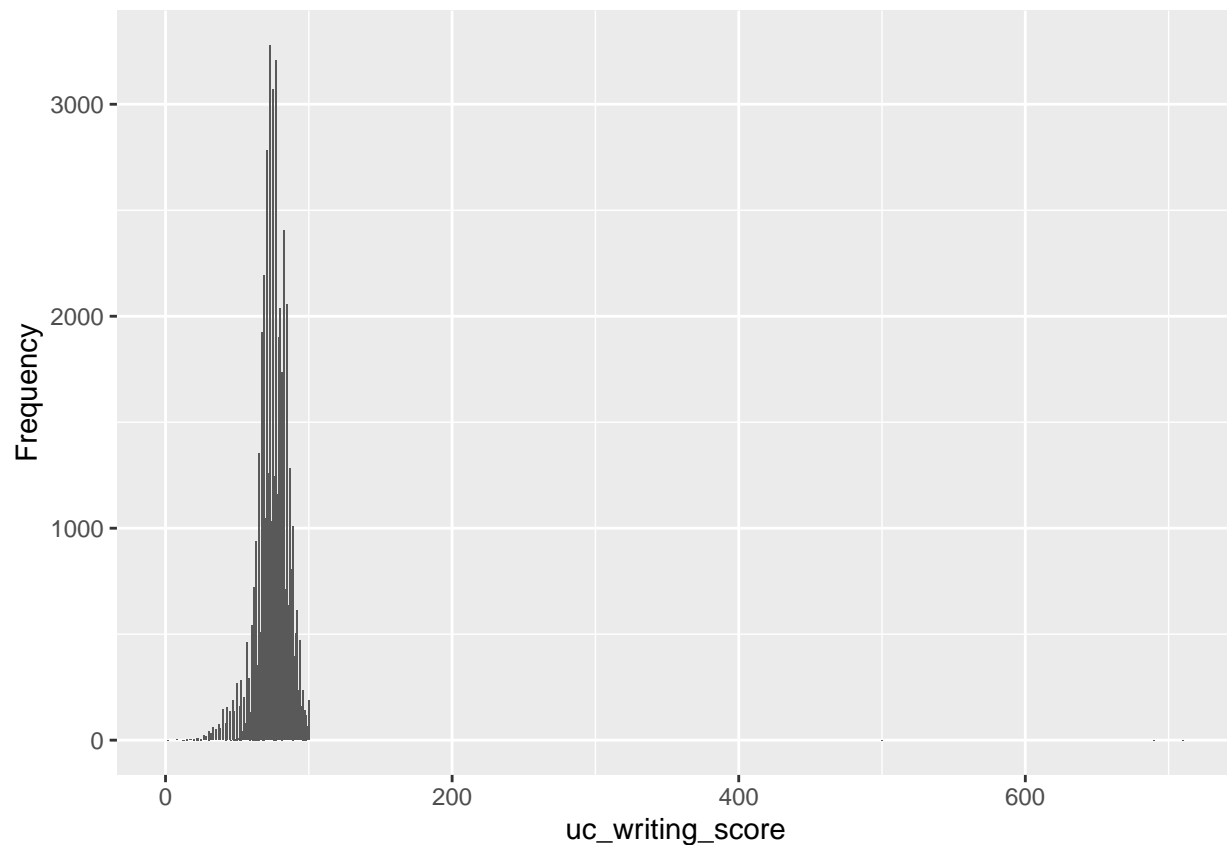
```
[1] -----
[1] Variable: uc_math_score, type: numeric, requested: yes
[1] Values (88 unique): NA, 83, 82, 98, 85, ...
[1] Missing: 48.9%
  Group.1 uc_math_score
1    2006    0.9993725
2    2007    0.9994105
3    2008    0.9972312
4    2009    0.9984897
5    2010    0.9983001
6    2011    0.2109065
7    2012    0.2032462
8    2013    0.2370147
9    2014    0.2458333
10   2015    0.2257943
11   2016    0.2360857
12   2017    0.2615005
13   2018    0.2596043
```

Warning: Removed 47426 rows containing non-finite values ('stat_count()').



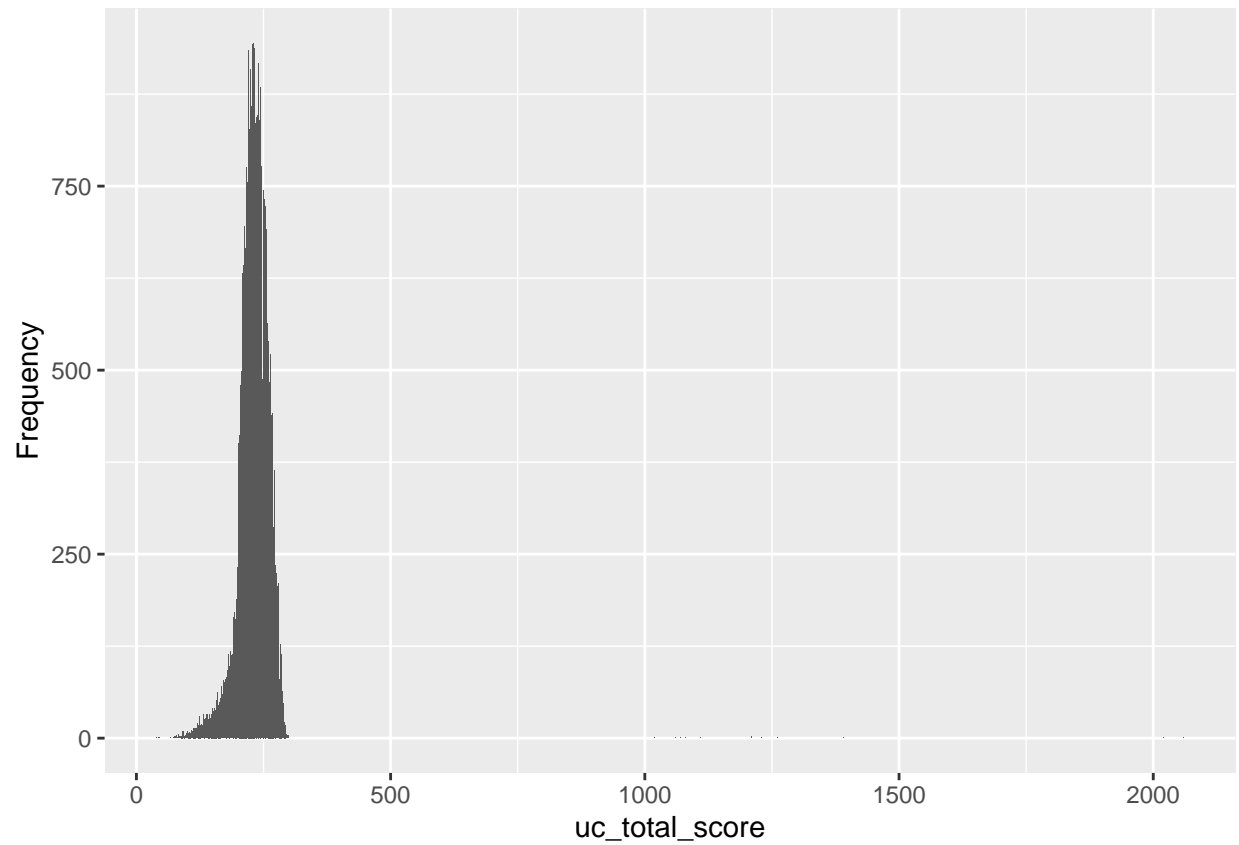
```
[1] -----
[1] Variable: uc_writing_score, type: numeric, requested: yes
[1] Values (82 unique): NA, 71, 88, 94, 80, ...
[1] Missing: 48.9%
Group.1 uc_writing_score
1      2006      0.9995294
2      2007      0.9994105
3      2008      0.9973850
4      2009      0.9984897
5      2010      0.9983001
6      2011      0.2114731
7      2012      0.2036771
8      2013      0.2378017
9      2014      0.2458333
10     2015      0.2257943
11     2016      0.2360857
12     2017      0.2615005
13     2018      0.2597193
```

Warning: Removed 47442 rows containing non-finite values ('stat_count()').

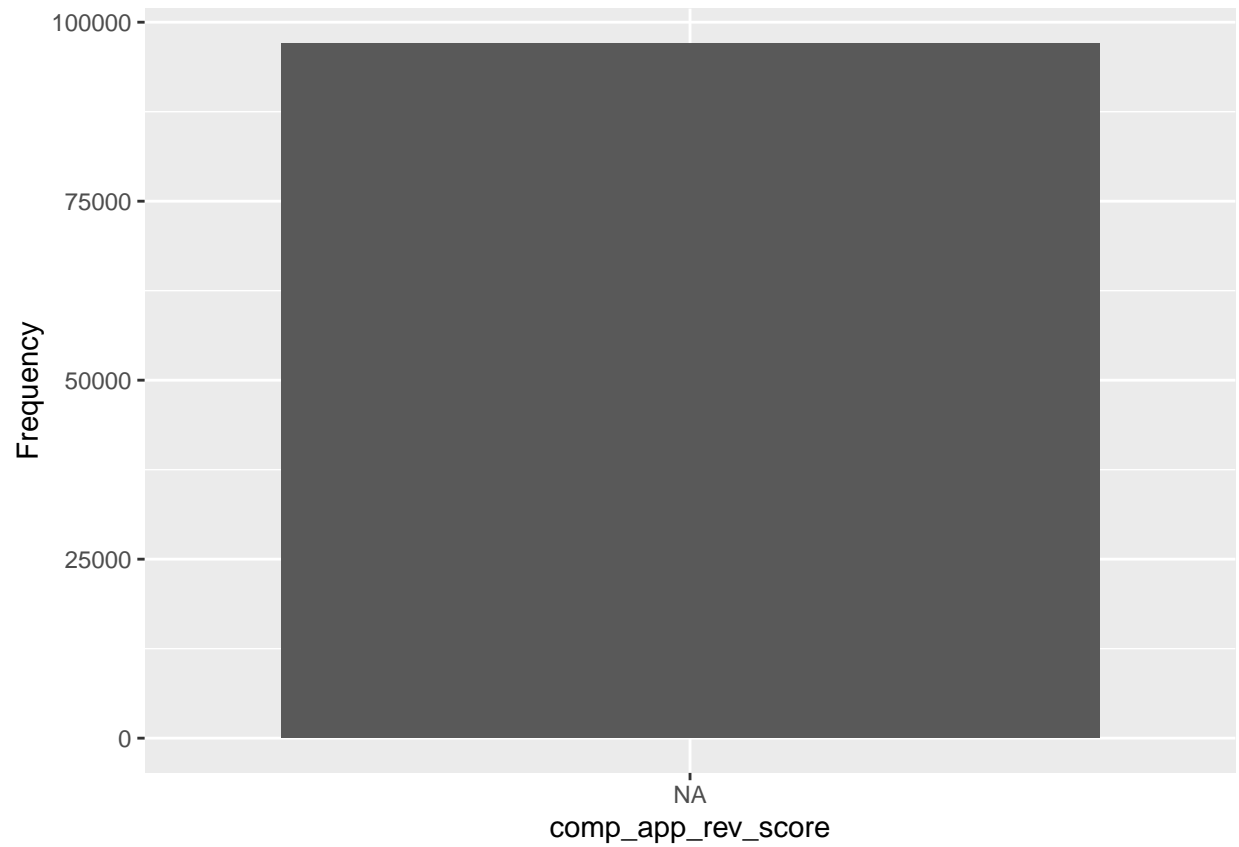


```
[1] -----
[1] Variable: uc_total_score, type: numeric, requested: yes
[1] Values (241 unique): NA, 224, 262, 280, 253, ...
[1] Missing: 48.9%
  Group.1 uc_total_score
1    2006      0.9993725
2    2007      0.9994105
3    2008      0.9972312
4    2009      0.9984897
5    2010      0.9983001
6    2011      0.2109065
7    2012      0.2032462
8    2013      0.2370147
9    2014      0.2458333
10   2015      0.2257943
11   2016      0.2360857
12   2017      0.2615005
13   2018      0.2597193
```

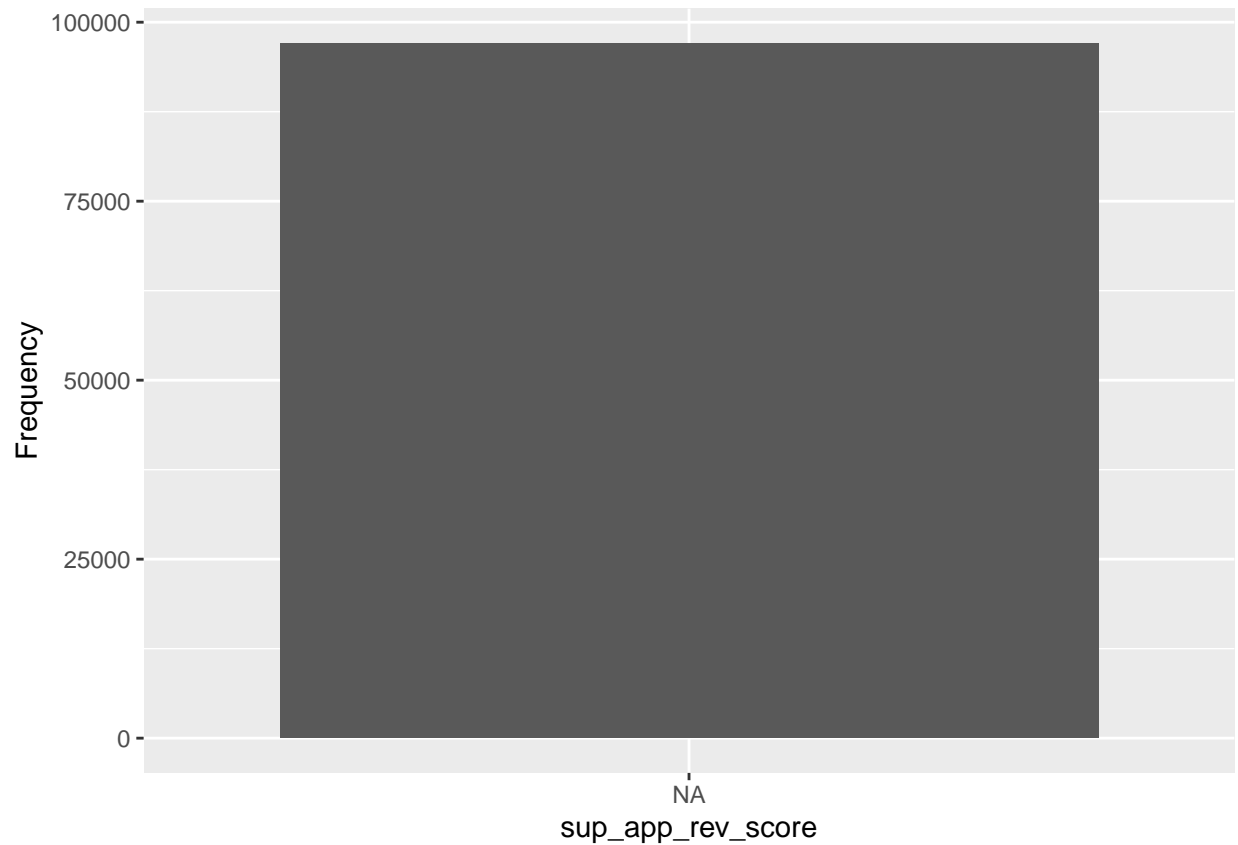
Warning: Removed 47427 rows containing non-finite values ('stat_count()').



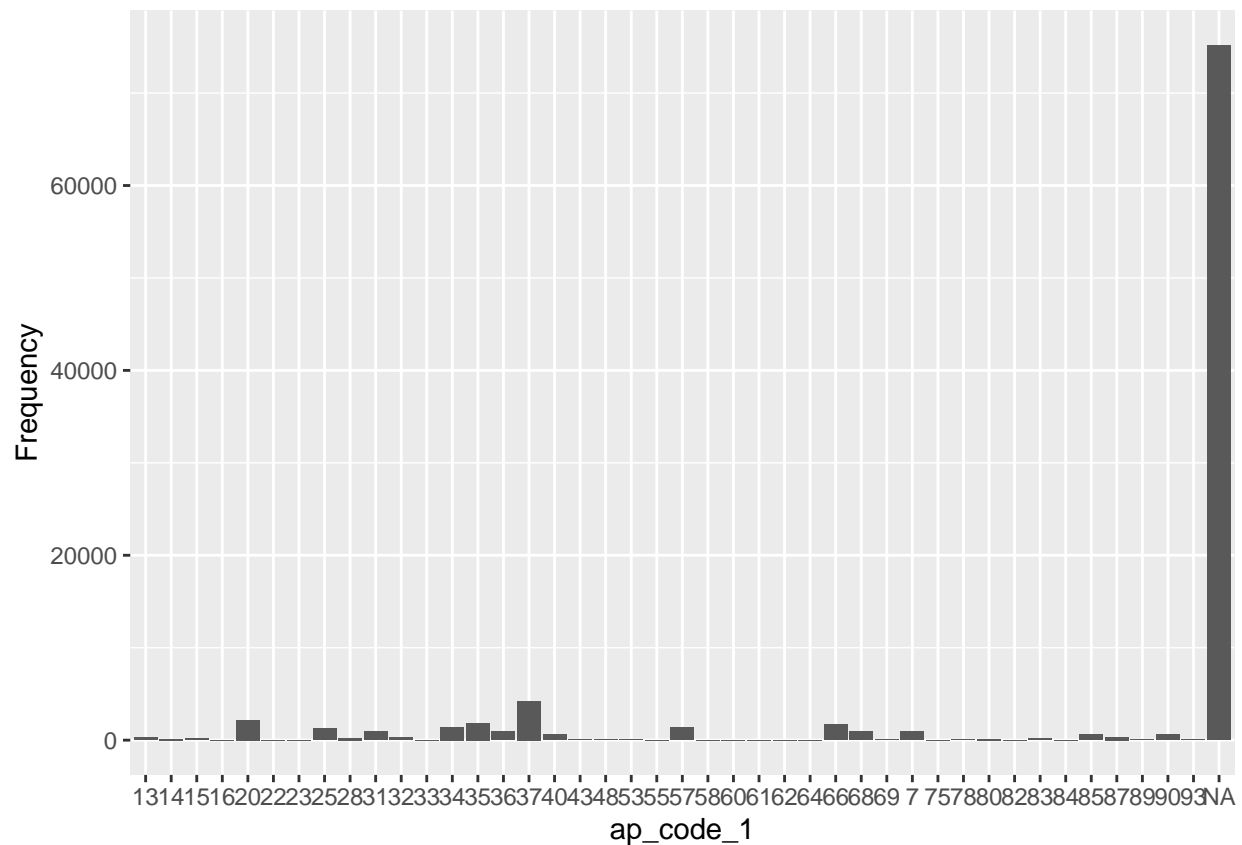
```
[1] -----  
[1] Variable: comp_app_rev_score, type: logical, requested: yes  
[1] Values (1 unique): NA  
[1] Missing: 100%
```



```
[1] -----  
[1] Variable: sup_app_rev_score, type: logical, requested: yes  
[1] Values (1 unique): NA  
[1] Missing: 100%
```

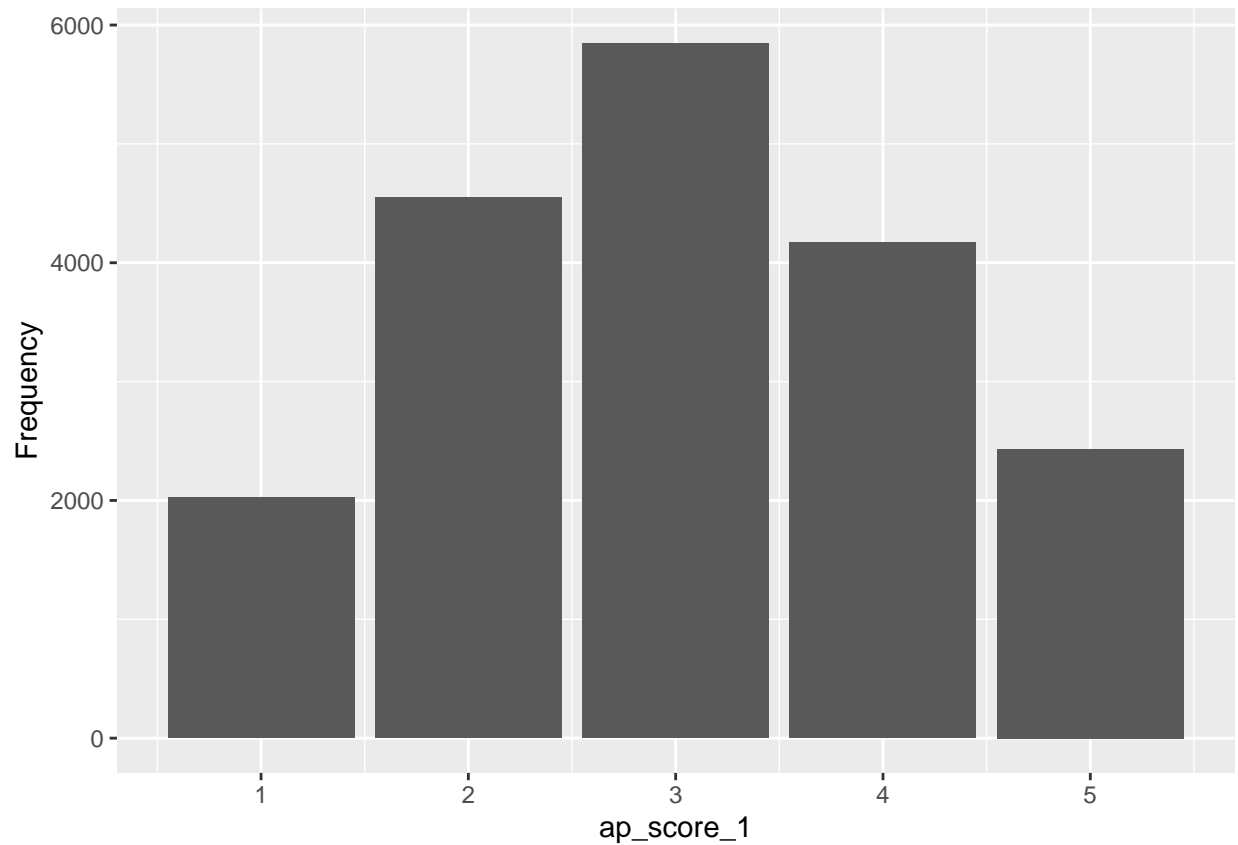



```
[1] -----
[1] Variable: ap_code_1, type: character, requested: yes
[1] Values (43 unique): NA, 66, 34, 35, 20, ...
[1] Missing: 77.4%
  Group.1 ap_code_1
1      2006 1.0000000
2      2007 0.9998526
3      2008 1.0000000
4      2009 1.0000000
5      2010 1.0000000
6      2011 0.9923513
7      2012 0.9918127
8      2013 0.9815058
9      2014 0.9085937
10     2015 0.5260968
11     2016 0.4491334
12     2017 0.3627663
13     2018 0.3256269
```



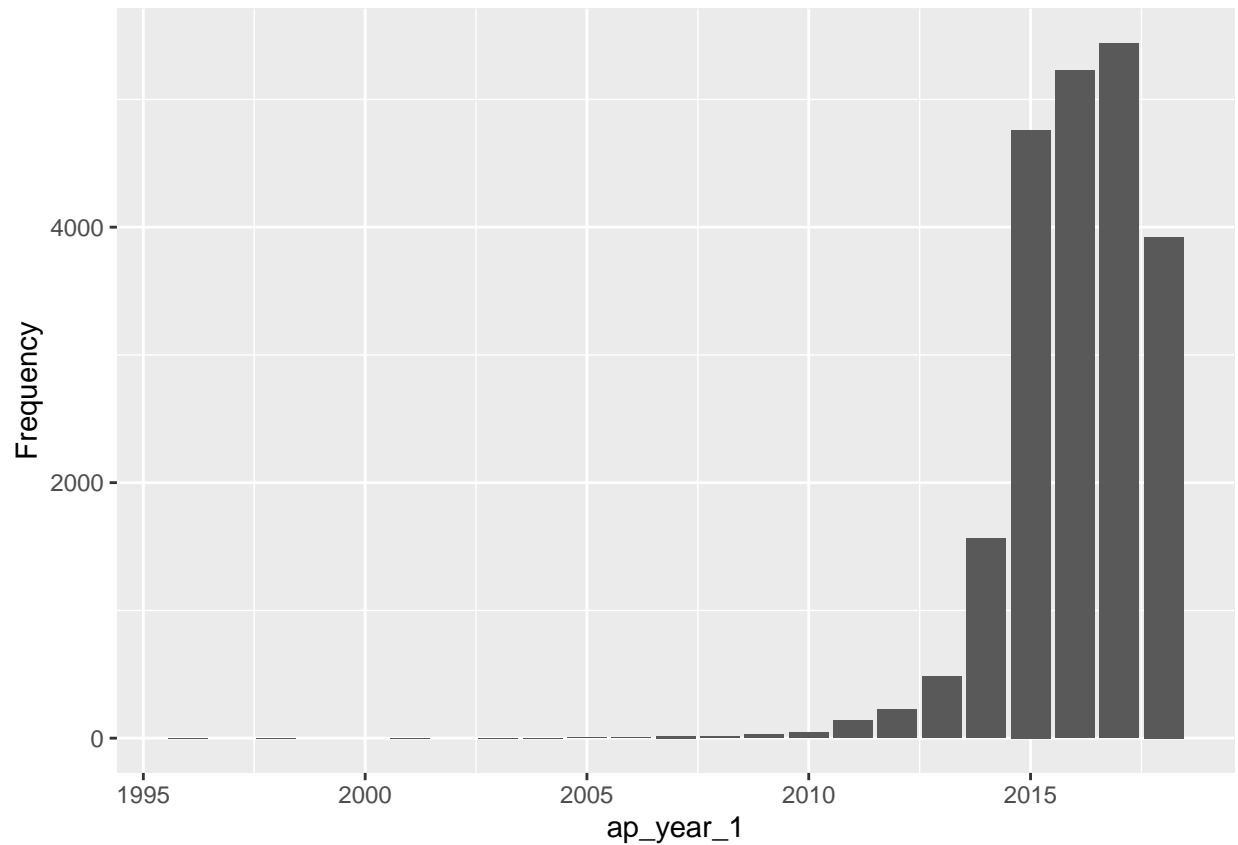
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_1, type: numeric, requested: yes
[1] Values (6 unique): NA, 4, 5, 2, 3, ...
[1] Missing: 80.4%
  Group.1 ap_score_1
1    2006  1.0000000
2    2007  0.9998526
3    2008  1.0000000
4    2009  1.0000000
5    2010  1.0000000
6    2011  0.9930595
7    2012  0.9928182
8    2013  0.9837356
9    2014  0.9161458
10   2015  0.5786687
11   2016  0.5245990
12   2017  0.4242050
13   2018  0.4469749
```

Warning: Removed 78020 rows containing non-finite values ('stat_count()').



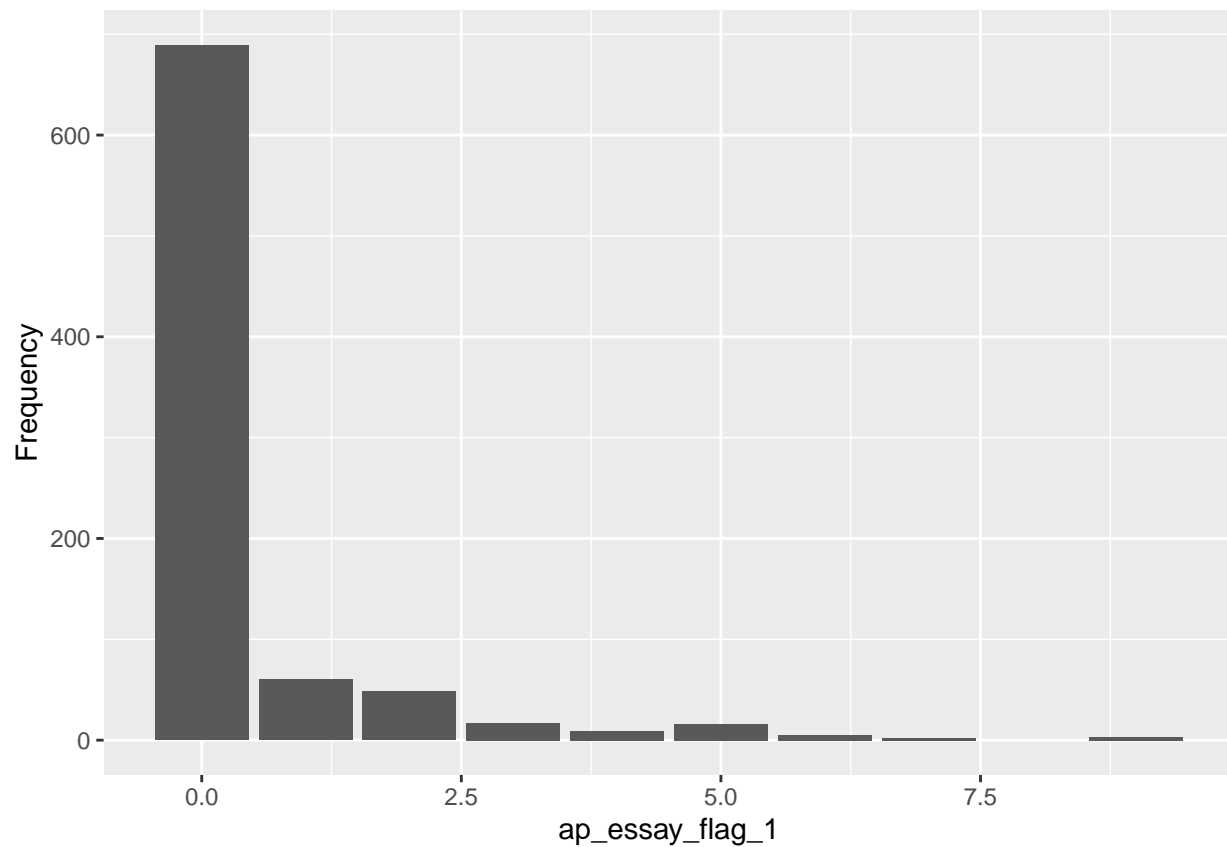
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_1, type: integer, requested: yes
[1] Values (20 unique): NA, 2017, 2018, 2016, 2015, ...
[1] Missing: 77.4%
  Group.1 ap_year_1
1      2006 1.0000000
2      2007 0.9998526
3      2008 1.0000000
4      2009 1.0000000
5      2010 1.0000000
6      2011 0.9923513
7      2012 0.9918127
8      2013 0.9815058
9      2014 0.9085937
10     2015 0.5260968
11     2016 0.4491334
12     2017 0.3627663
13     2018 0.3256269
```

Warning: Removed 75163 rows containing non-finite values ('stat_count()').

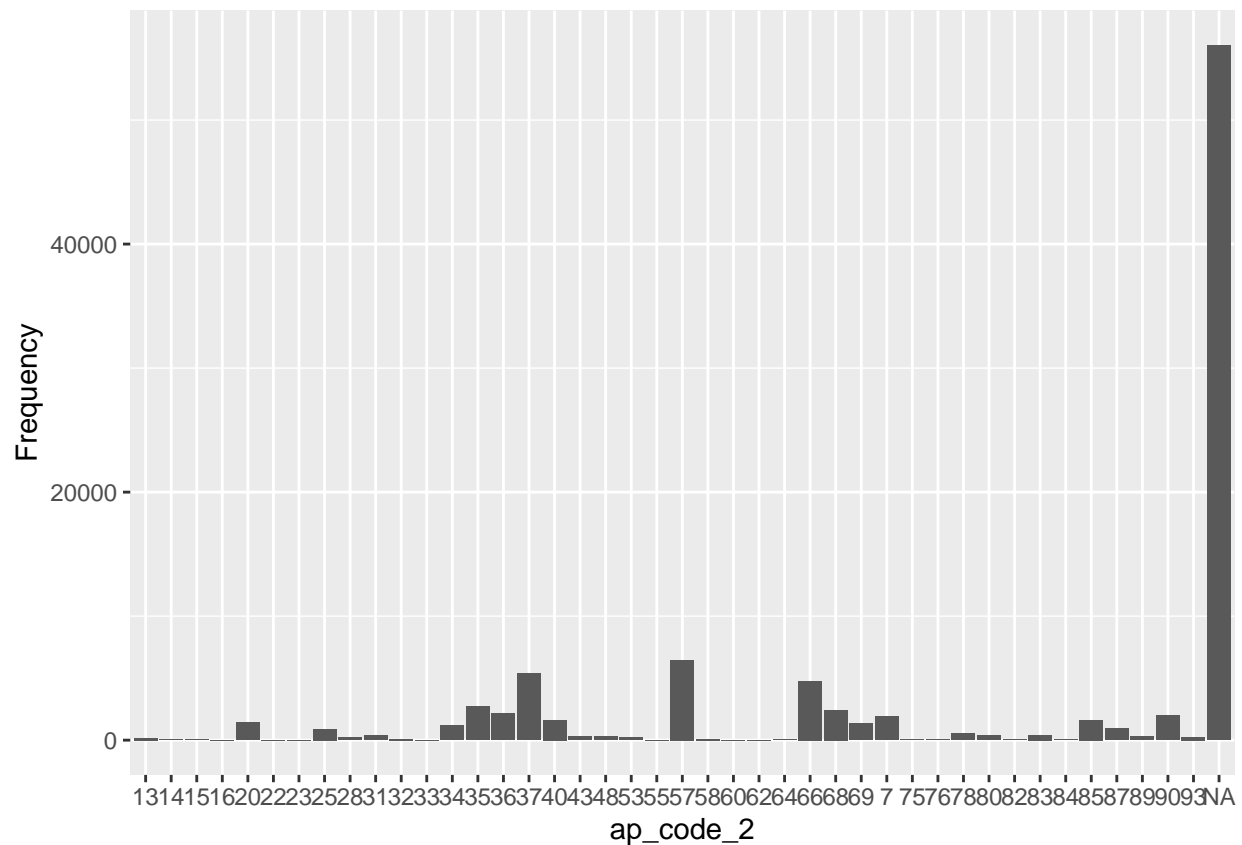


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_1, type: numeric, requested: yes
[1] Values (10 unique): NA, 0, 1, 3, 2, ...
[1] Missing: 99.1%
```

Warning: Removed 96200 rows containing non-finite values ('stat_count()').

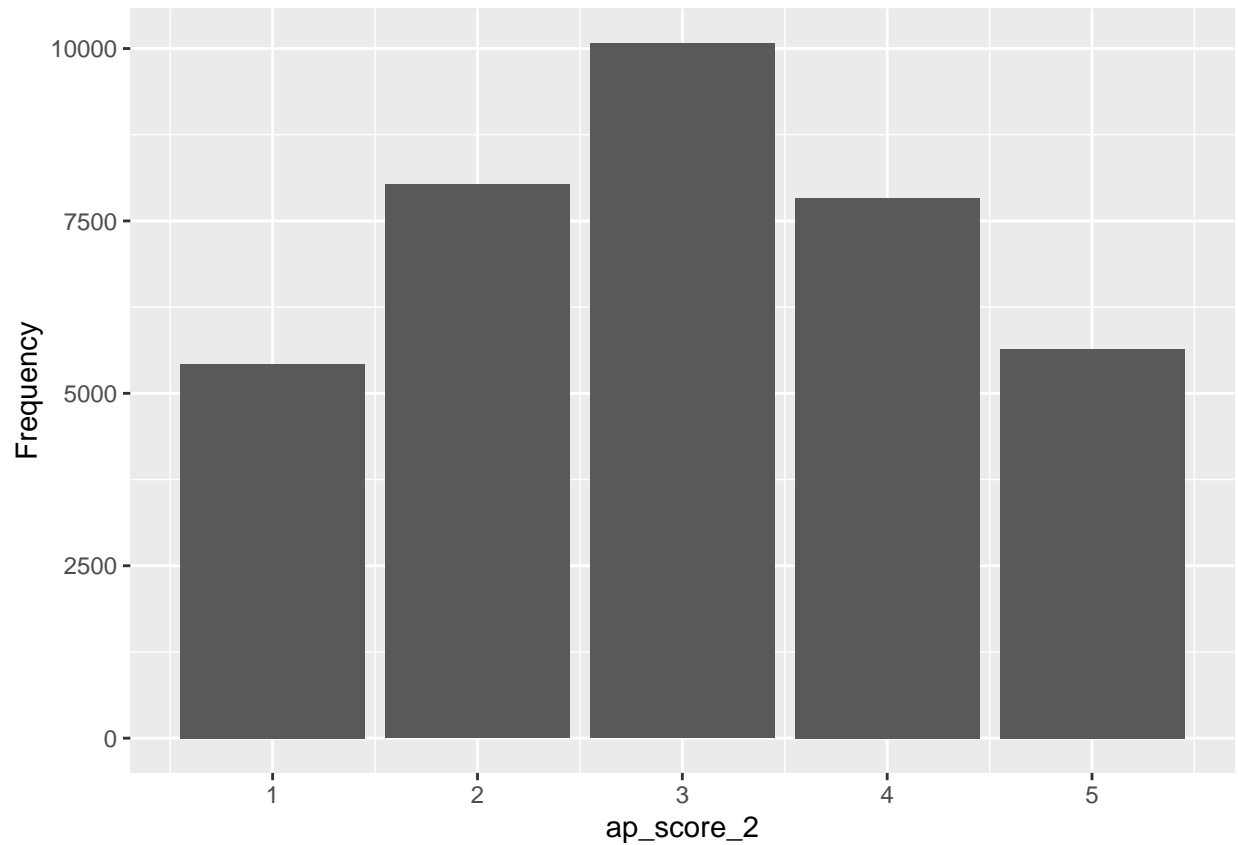


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_2, type: character, requested: yes
[1] Values (43 unique): NA, 80, 57, 83, 37, ...
[1] Missing: 57.7%
  Group.1 ap_code_2
1      2006 0.9995294
2      2007 0.9988209
3      2008 0.9972312
4      2009 0.9981541
5      2010 0.9983001
6      2011 0.3270538
7      2012 0.3761850
8      2013 0.3610965
9      2014 0.3757813
10     2015 0.3765759
11     2016 0.3837873
12     2017 0.3660595
13     2018 0.3806073
```



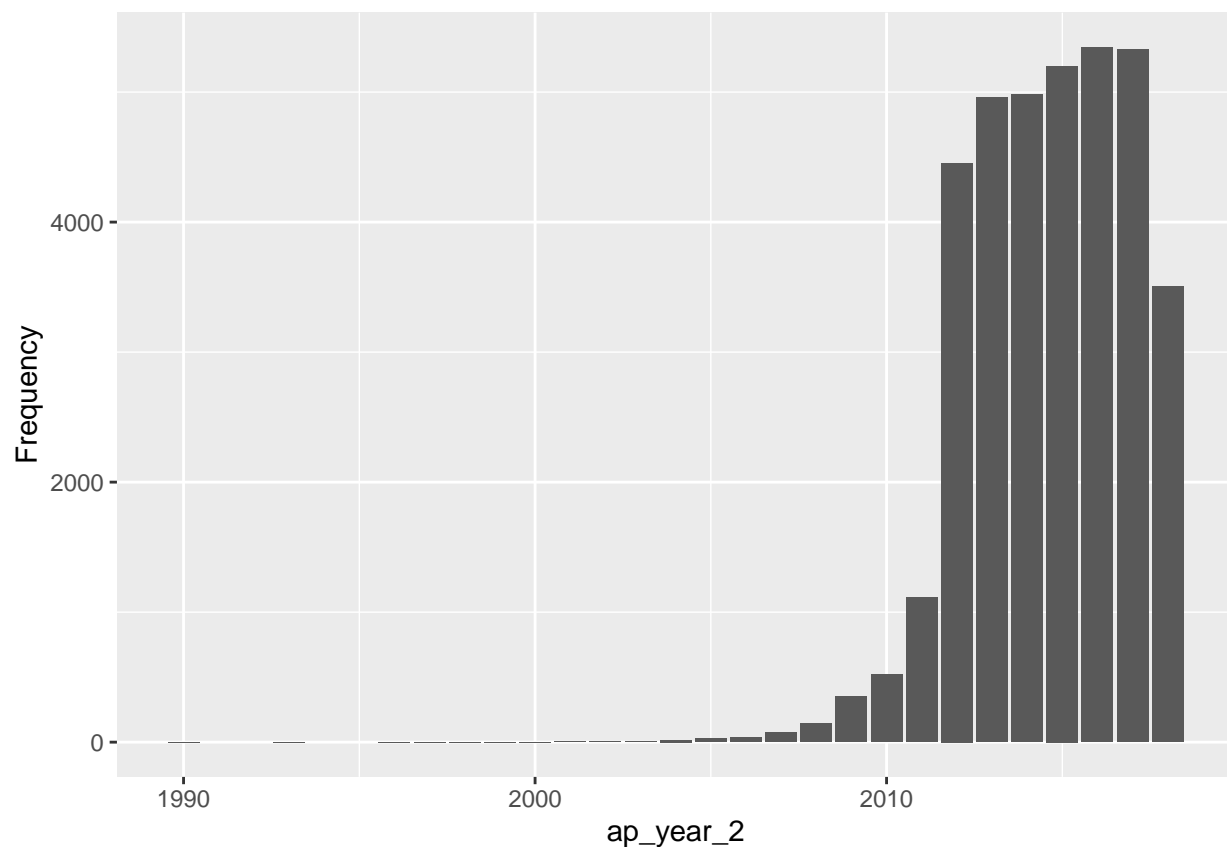
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_2, type: numeric, requested: yes
[1] Values (6 unique): NA, 4, 5, 3, 1, ...
[1] Missing: 61.9%
  Group.1 ap_score_2
1    2006  0.9995294
2    2007  0.9988209
3    2008  0.9973850
4    2009  0.9981541
5    2010  0.9986092
6    2011  0.3783286
7    2012  0.4317725
8    2013  0.3991343
9    2014  0.4143229
10   2015  0.4331820
11   2016  0.4614060
12   2017  0.4256458
13   2018  0.4873476
```

Warning: Removed 60040 rows containing non-finite values ('stat_count()').



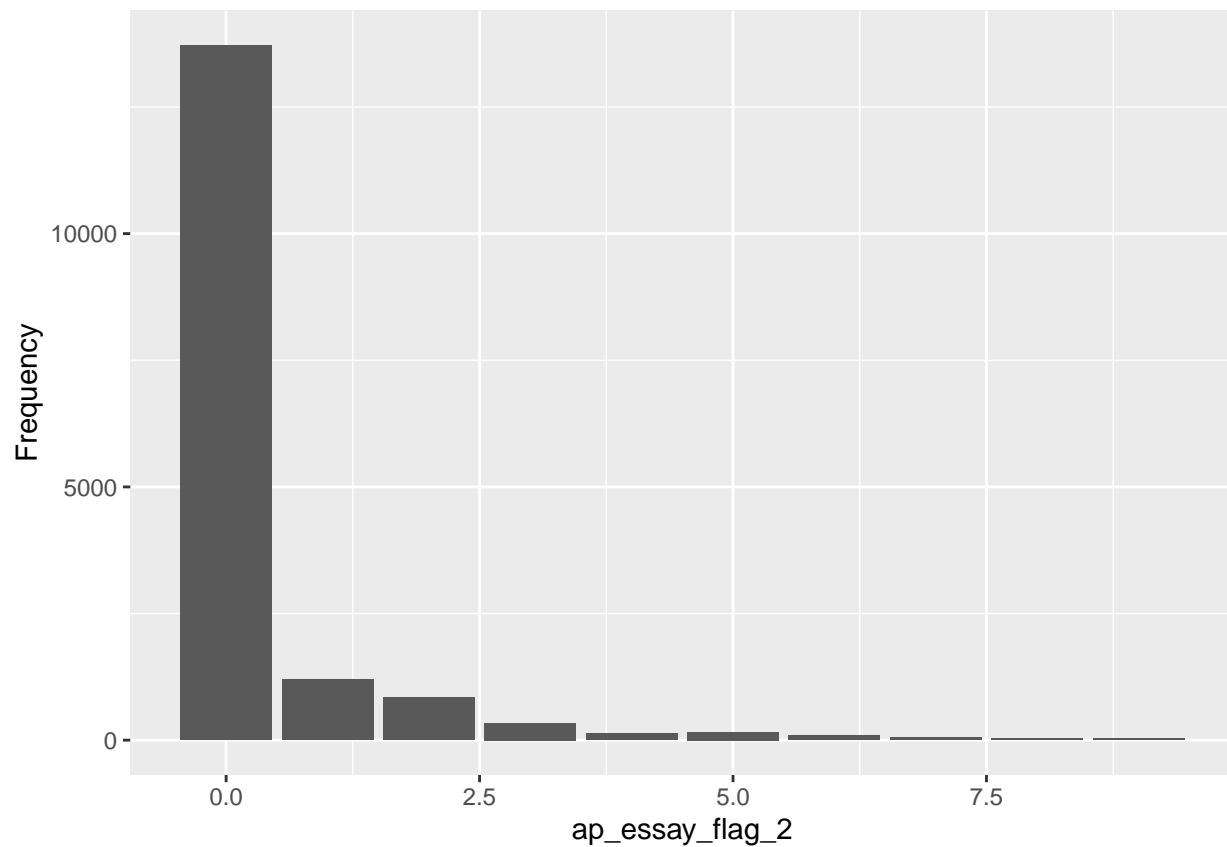
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_2, type: integer, requested: yes
[1] Values (26 unique): NA, 2017, 2018, 2016, 2015, ...
[1] Missing: 62.8%
Group.1 ap_year_2
1      2006 0.9995294
2      2007 0.9991157
3      2008 0.9981541
4      2009 0.9984897
5      2010 0.9984546
6      2011 0.9991501
7      2012 0.3836541
8      2013 0.3786726
9      2014 0.3766927
10     2015 0.3768280
11     2016 0.3837873
12     2017 0.3660595
13     2018 0.3806073
```

Warning: Removed 60976 rows containing non-finite values ('stat_count()').

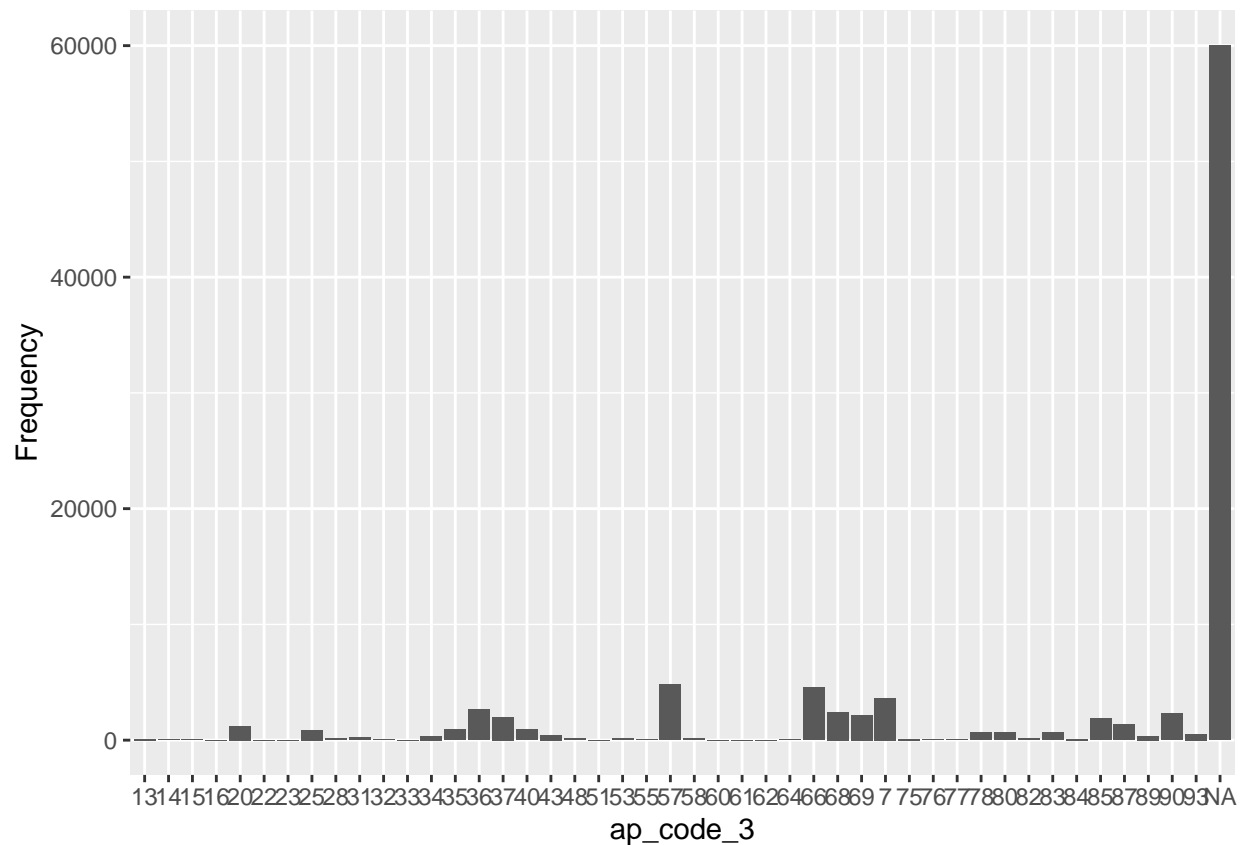


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_2, type: numeric, requested: yes
[1] Values (11 unique): NA, 0, 1, 4, 2, ...
[1] Missing: 82.9%
  Group.1 ap_essay_flag_2
1    2006      0.9996863
2    2007      0.9997052
3    2008      0.9981541
4    2009      0.9986575
5    2010      0.9990728
6    2011      0.4099150
7    2012      0.4671072
8    2013      0.4266789
9    2014      0.4438802
10   2015      0.9982350
11   2016      0.9981699
12   2017      0.9993825
13   2018      0.9998850
```

Warning: Removed 80463 rows containing non-finite values ('stat_count()').

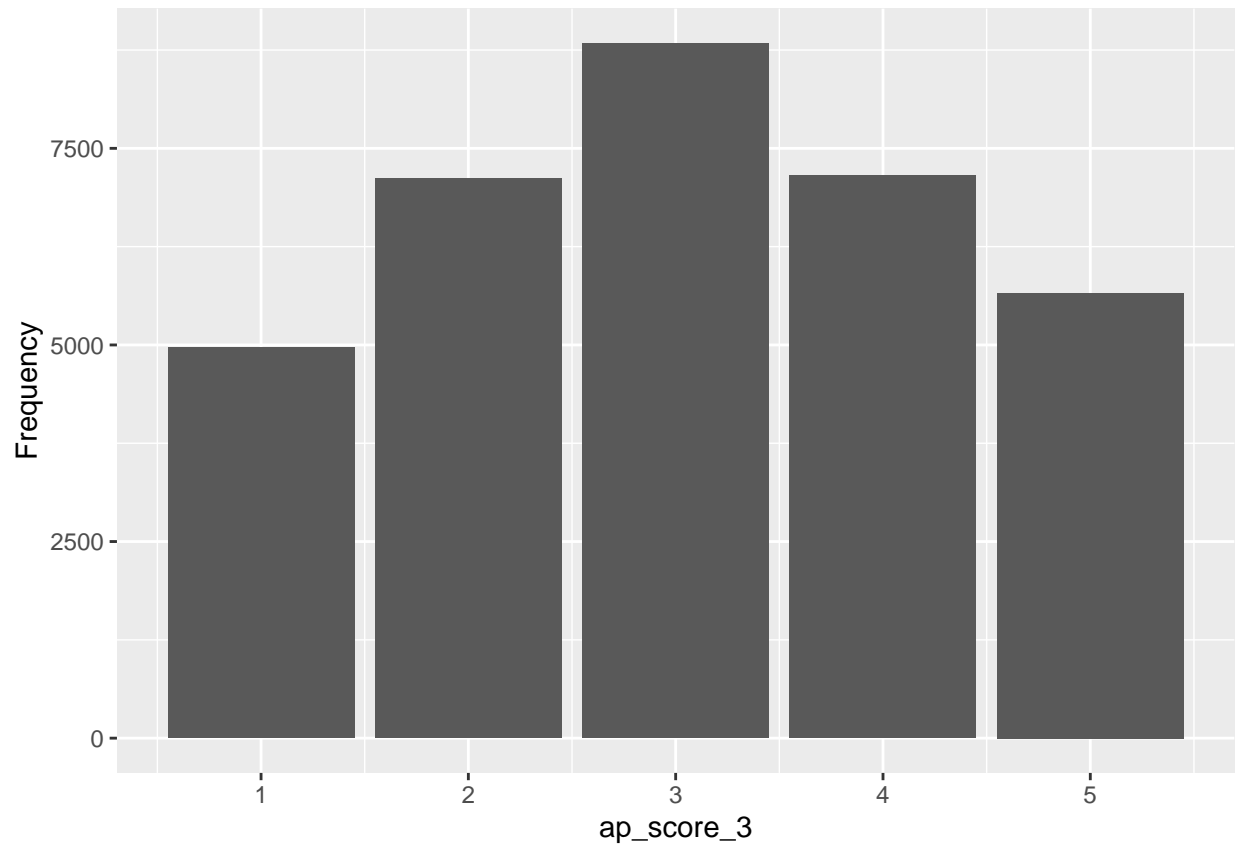


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_3, type: character, requested: yes
[1] Values (46 unique): NA, 80, 90, 66, 25, ...
[1] Missing: 61.9%
  Group.1 ap_code_3
1      2006 0.9995294
2      2007 0.9988209
3      2008 0.9976927
4      2009 0.9981541
5      2010 0.9987637
6      2011 0.4025496
7      2012 0.4478598
8      2013 0.4252361
9      2014 0.4291667
10     2015 0.4357035
11     2016 0.4385833
12     2017 0.4246167
13     2018 0.4401886
```



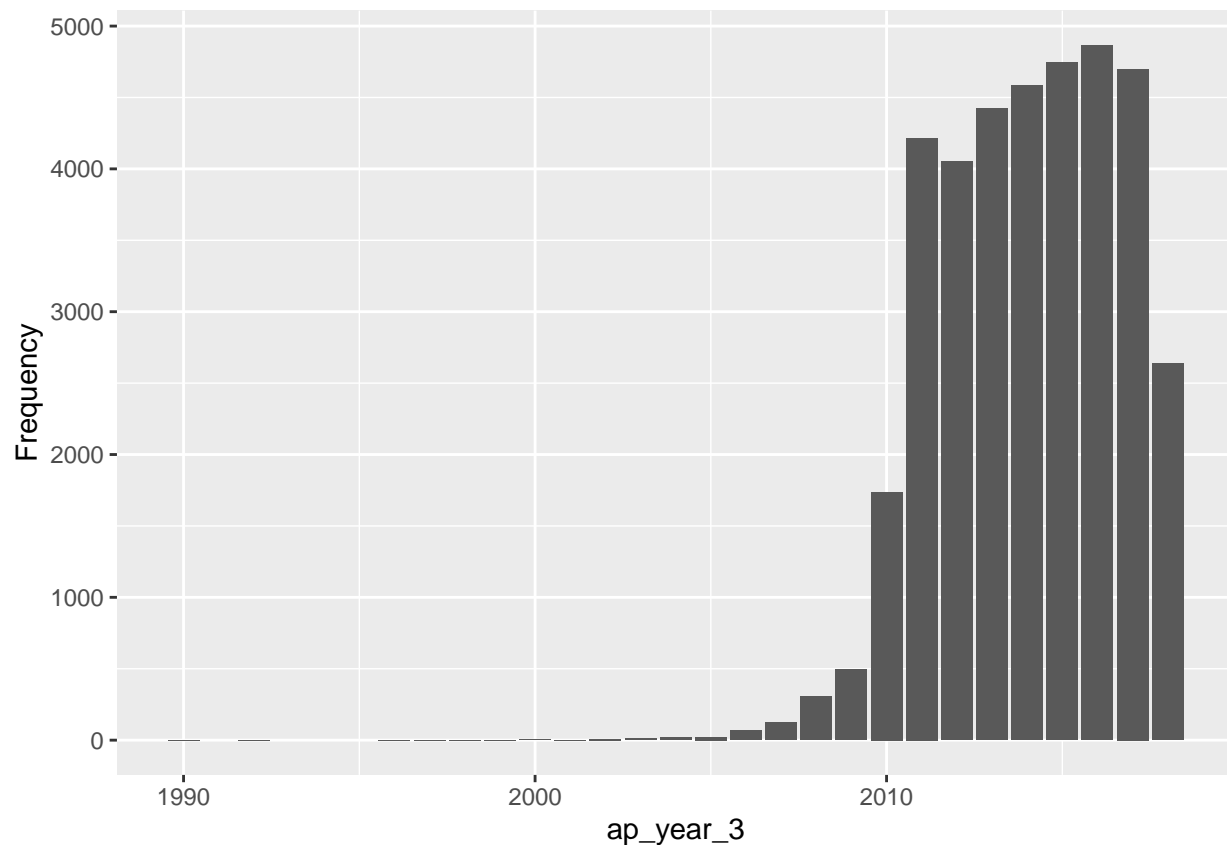
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_3, type: numeric, requested: yes
[1] Values (6 unique): NA, 5, 2, 3, 1, ...
[1] Missing: 65.2%
  Group.1 ap_score_3
1    2006 0.9995294
2    2007 0.9988209
3    2008 0.9976927
4    2009 0.9983219
5    2010 0.9990728
6    2011 0.4485836
7    2012 0.4903763
8    2013 0.4614376
9    2014 0.4631510
10   2015 0.4793243
11   2016 0.5006998
12   2017 0.4740146
13   2018 0.5232344
```

Warning: Removed 63313 rows containing non-finite values ('stat_count()').



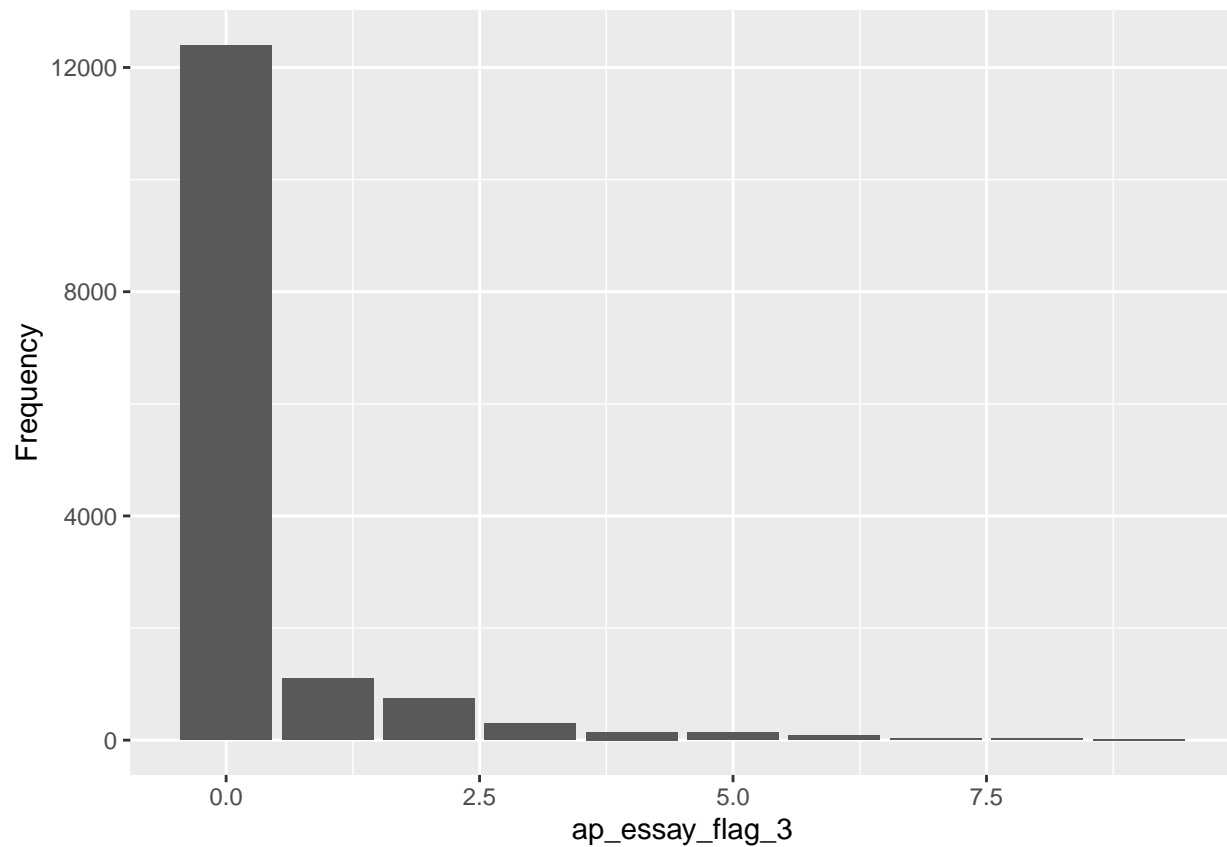
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_3, type: integer, requested: yes
[1] Values (26 unique): NA, 2018, 2017, 2015, 2016, ...
[1] Missing: 61.9%
  Group.1 ap_year_3
1      2006 0.9995294
2      2007 0.9988209
3      2008 0.9976927
4      2009 0.9981541
5      2010 0.9987637
6      2011 0.4025496
7      2012 0.4478598
8      2013 0.4252361
9      2014 0.4291667
10     2015 0.4357035
11     2016 0.4385833
12     2017 0.4246167
13     2018 0.4401886
```

Warning: Removed 60027 rows containing non-finite values ('stat_count()').



```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_3, type: numeric, requested: yes
[1] Values (11 unique): NA, 0, 2, 1, 4, ...
[1] Missing: 84.6%
  Group.1 ap_essay_flag_3
1    2006      0.9996863
2    2007      0.9997052
3    2008      0.9984618
4    2009      0.9986575
5    2010      0.9993819
6    2011      0.4743626
7    2012      0.5193910
8    2013      0.4845226
9    2014      0.4871094
10   2015      0.9981089
11   2016      0.9986005
12   2017      0.9994854
13   2018      0.9998850
```

Warning: Removed 82063 rows containing non-finite values ('stat_count()').



[1] is used in feature engineering

[1] -----

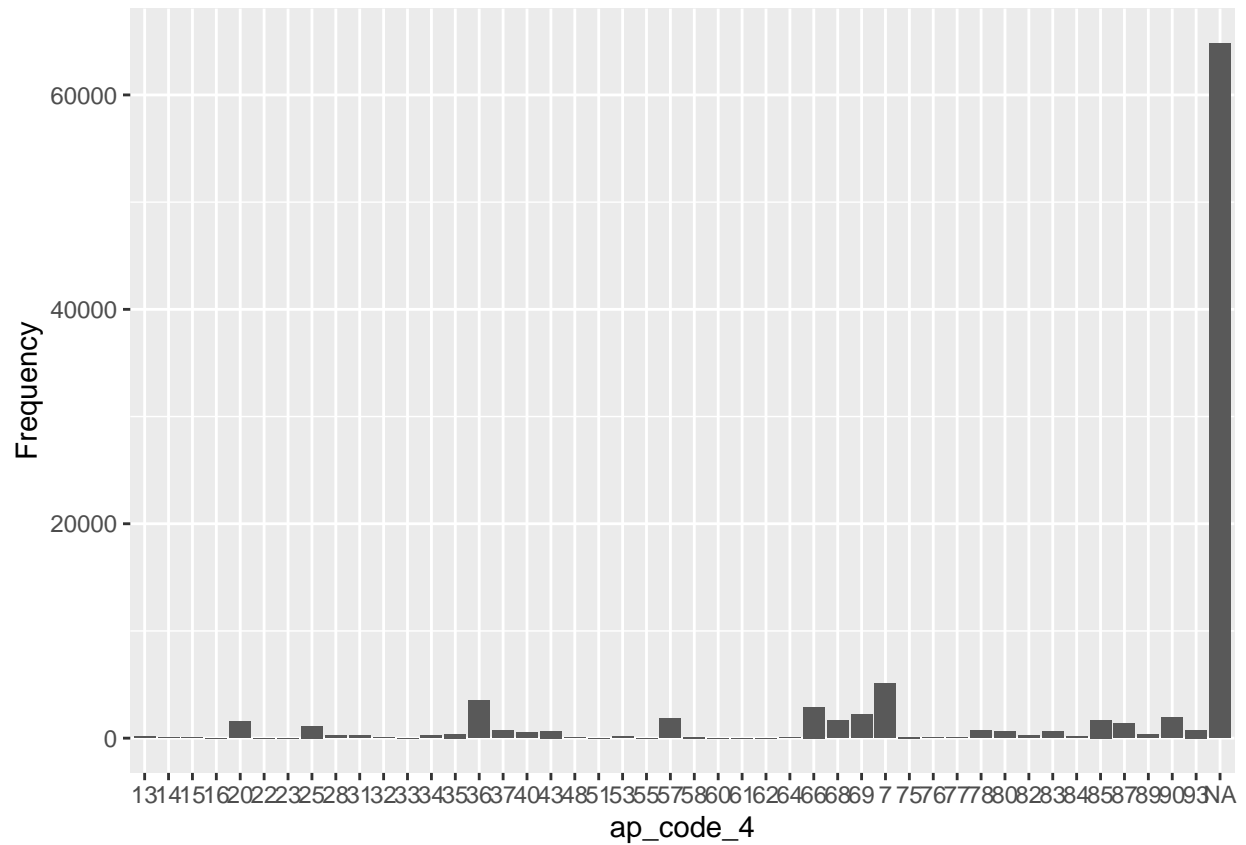
[1] Variable: ap_code_4, type: character, requested: yes

[1] Values (46 unique): NA, 68, 36, 89, 43, ...

[1] Missing: 66.8%

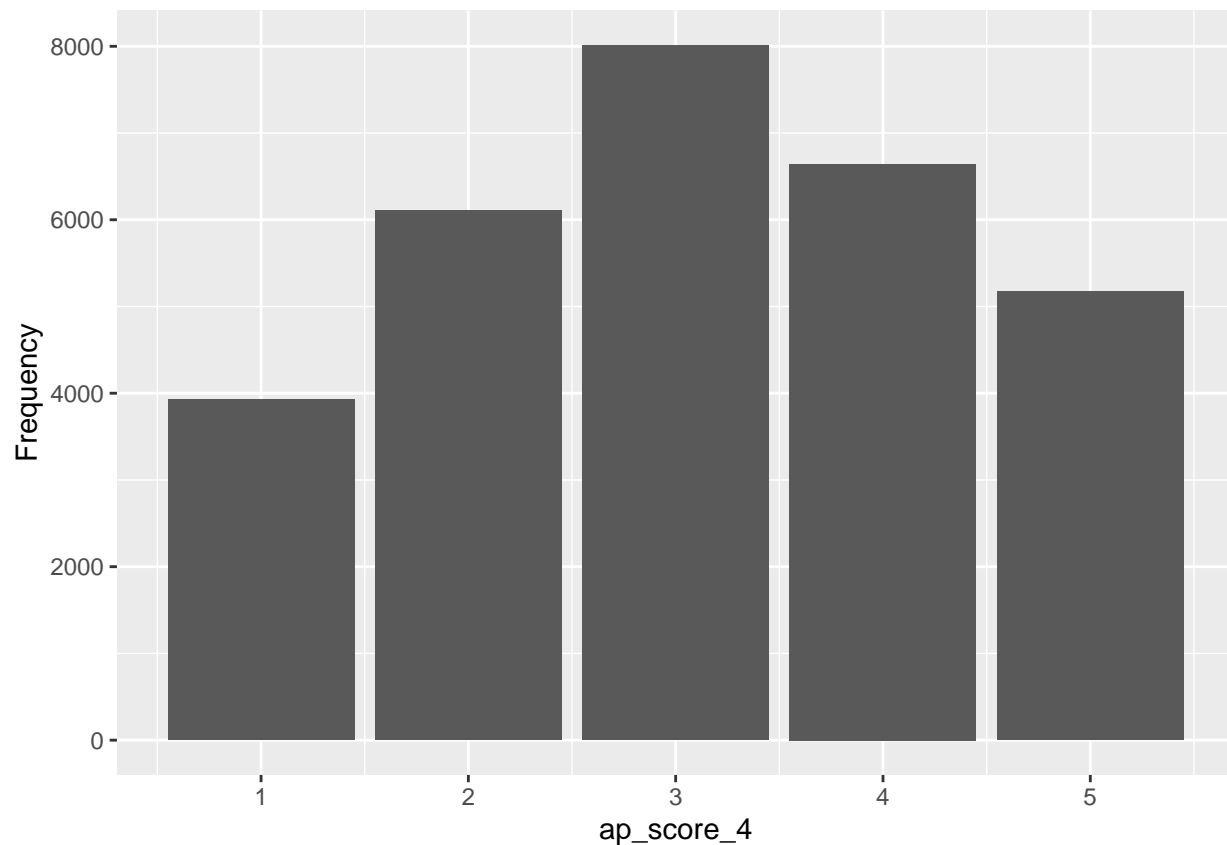
Group.1 ap_code_4

1	2006	0.9995294
2	2007	0.9988209
3	2008	0.9984618
4	2009	0.9983219
5	2010	0.9989183
6	2011	0.4864023
7	2012	0.5277219
8	2013	0.5019675
9	2014	0.4968750
10	2015	0.5065557
11	2016	0.5073743
12	2017	0.4943913
13	2018	0.5144928



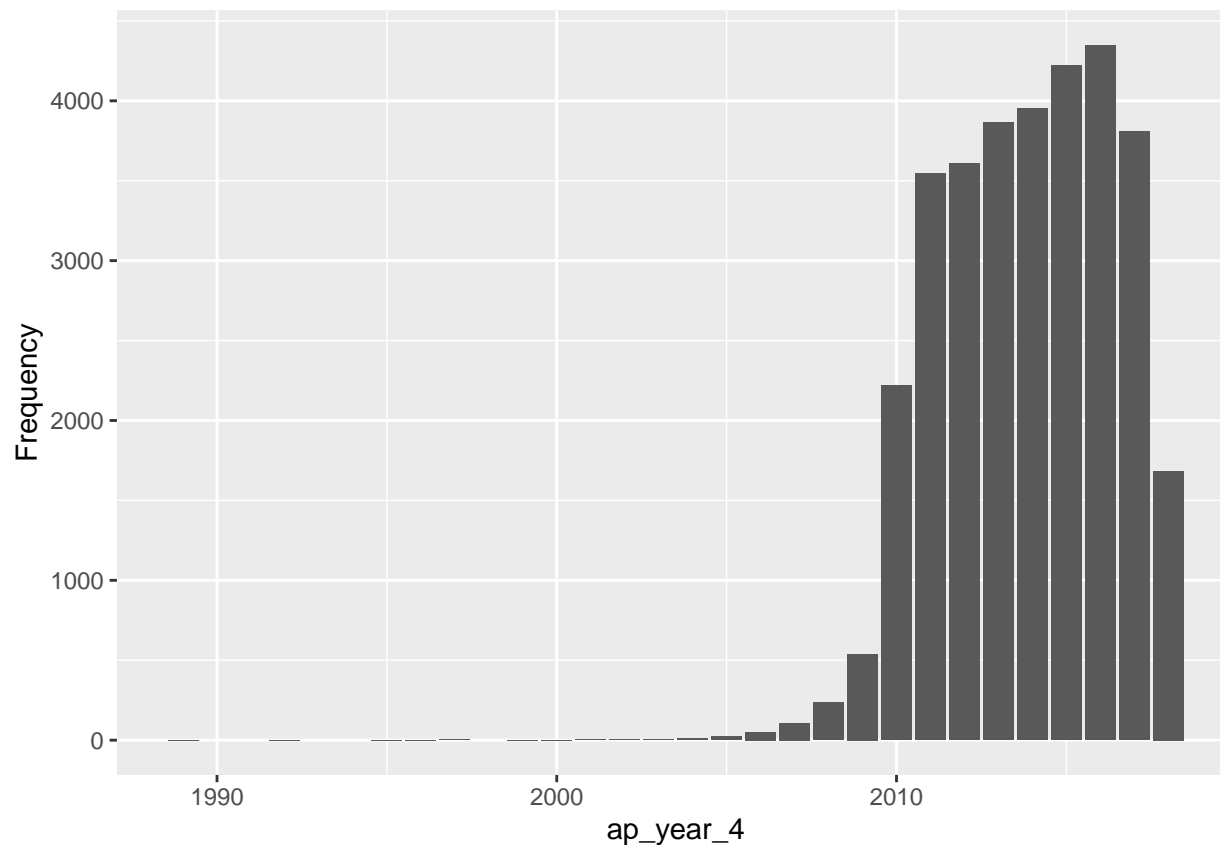
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_4, type: numeric, requested: yes
[1] Values (6 unique): NA, 3, 5, 4, 2, ...
[1] Missing: 69.2%
[1] Most missing: 2006 100%, Least missing: 2011 52.2%
```

Warning: Removed 67182 rows containing non-finite values ('stat_count()').



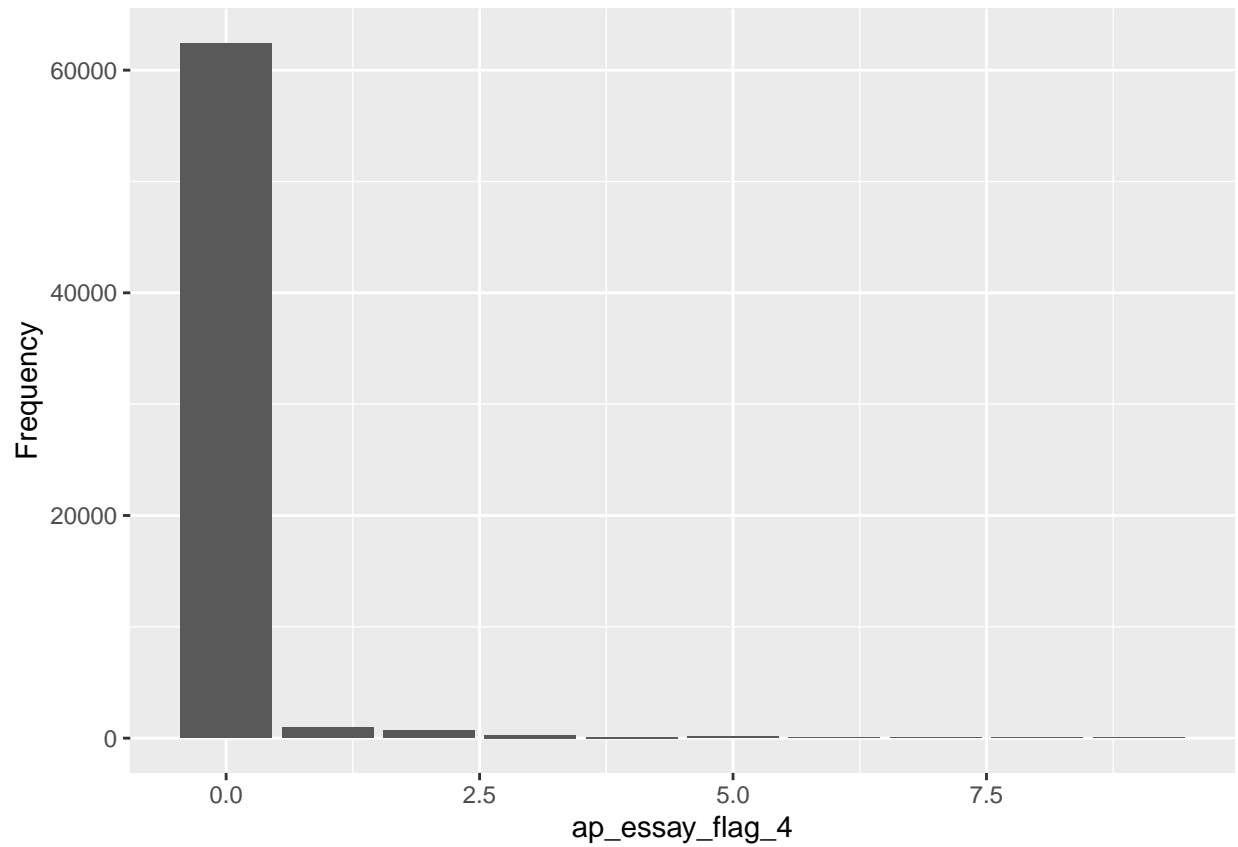
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_4, type: integer, requested: yes
[1] Values (26 unique): NA, 2018, 2017, 2016, 2014, ...
[1] Missing: 66.8%
  Group.1 ap_year_4
1      2006 0.9995294
2      2007 0.9988209
3      2008 0.9984618
4      2009 0.9983219
5      2010 0.9989183
6      2011 0.4864023
7      2012 0.5277219
8      2013 0.5019675
9      2014 0.4968750
10     2015 0.5065557
11     2016 0.5073743
12     2017 0.4943913
13     2018 0.5144928
```

Warning: Removed 64812 rows containing non-finite values ('stat_count()').

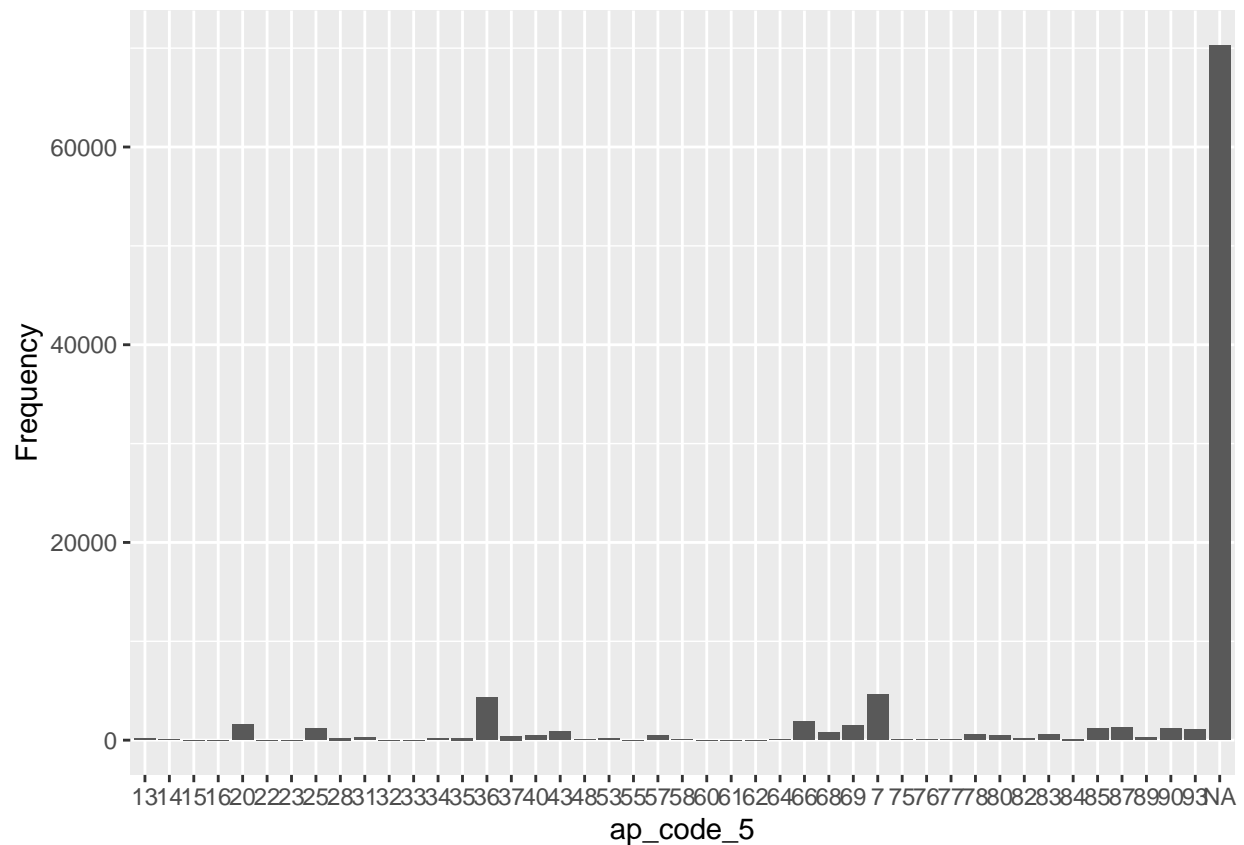


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_4, type: numeric, requested: yes
[1] Values (11 unique): 0, NA, 2, 6, 1, ...
[1] Missing: 33.2%
  Group.1 ap_essay_flag_4
1    2006    0.9990588235
2    2007    0.9985261606
3    2008    0.9955391478
4    2009    0.9974827991
5    2010    0.9970638232
6    2011    0.0004249292
7    2012    0.0005745475
8    2013    0.0003934942
9    2014    0.0037760417
10   2015    0.0037821483
11   2016    0.0055980192
12   2017    0.0028815478
13   2018    0.0113871636
```

Warning: Removed 32260 rows containing non-finite values ('stat_count()').

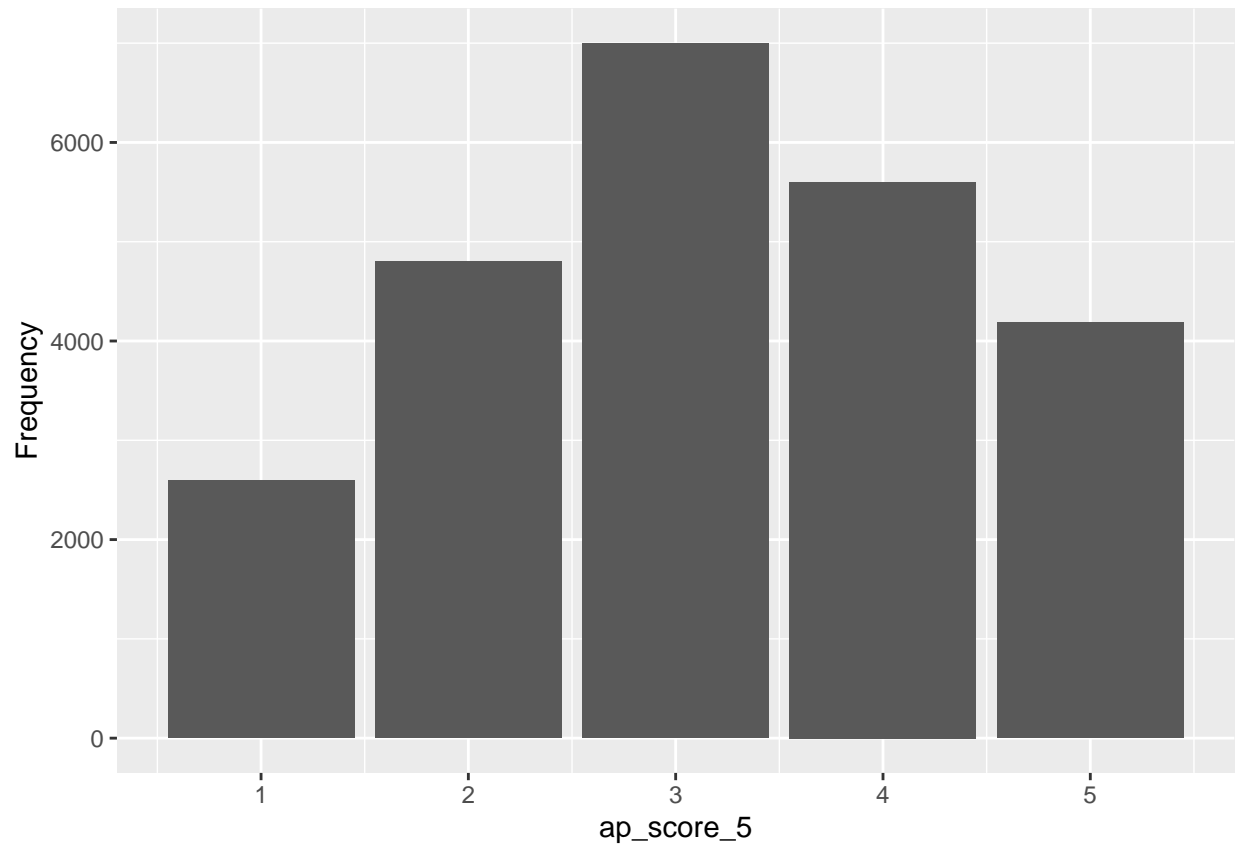


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_5, type: character, requested: yes
[1] Values (45 unique): NA, 7, 68, 57, 36, ...
[1] Missing: 72.4%
[1] Most missing: 2006 100%, Least missing: 2017 57.5%
```



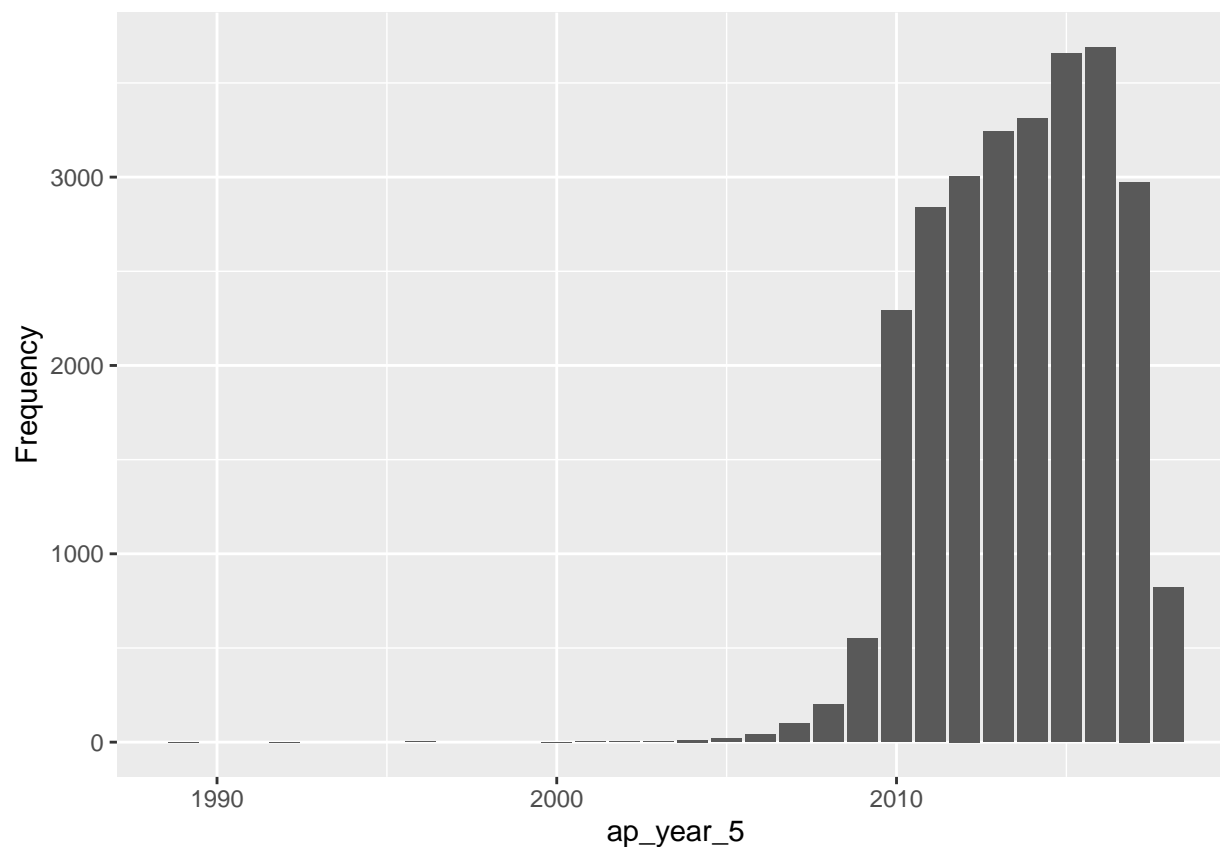
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_5, type: numeric, requested: yes
[1] Values (6 unique): NA, 3, 5, 4, 2, ...
[1] Missing: 75.1%
[1] Most missing: 2006 100%, Least missing: 2017 59.5%
```

Warning: Removed 72868 rows containing non-finite values ('stat_count()').



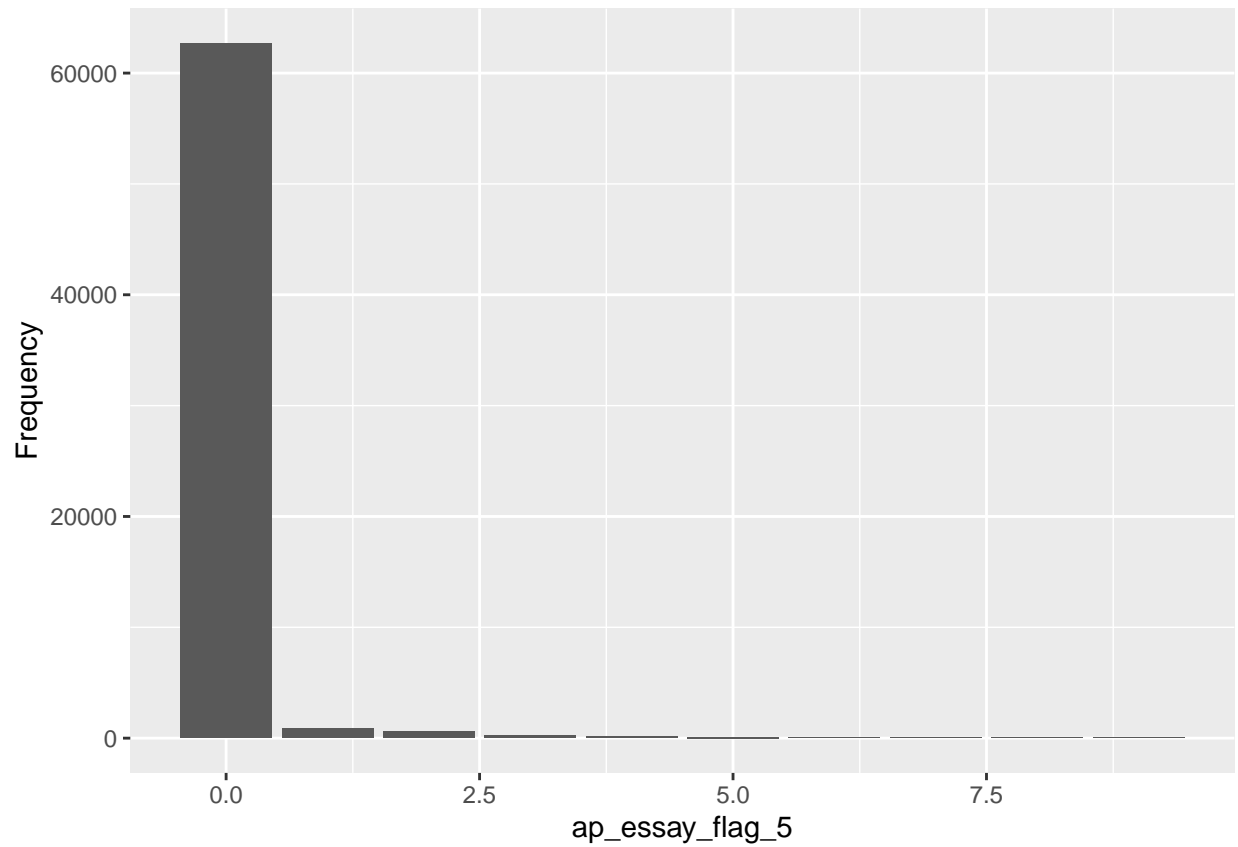
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_5, type: integer, requested: yes
[1] Values (23 unique): NA, 2017, 2018, 2016, 2015, ...
[1] Missing: 72.4%
[1] Most missing: 2006 100%, Least missing: 2017 57.5%
```

Warning: Removed 70285 rows containing non-finite values ('stat_count()').

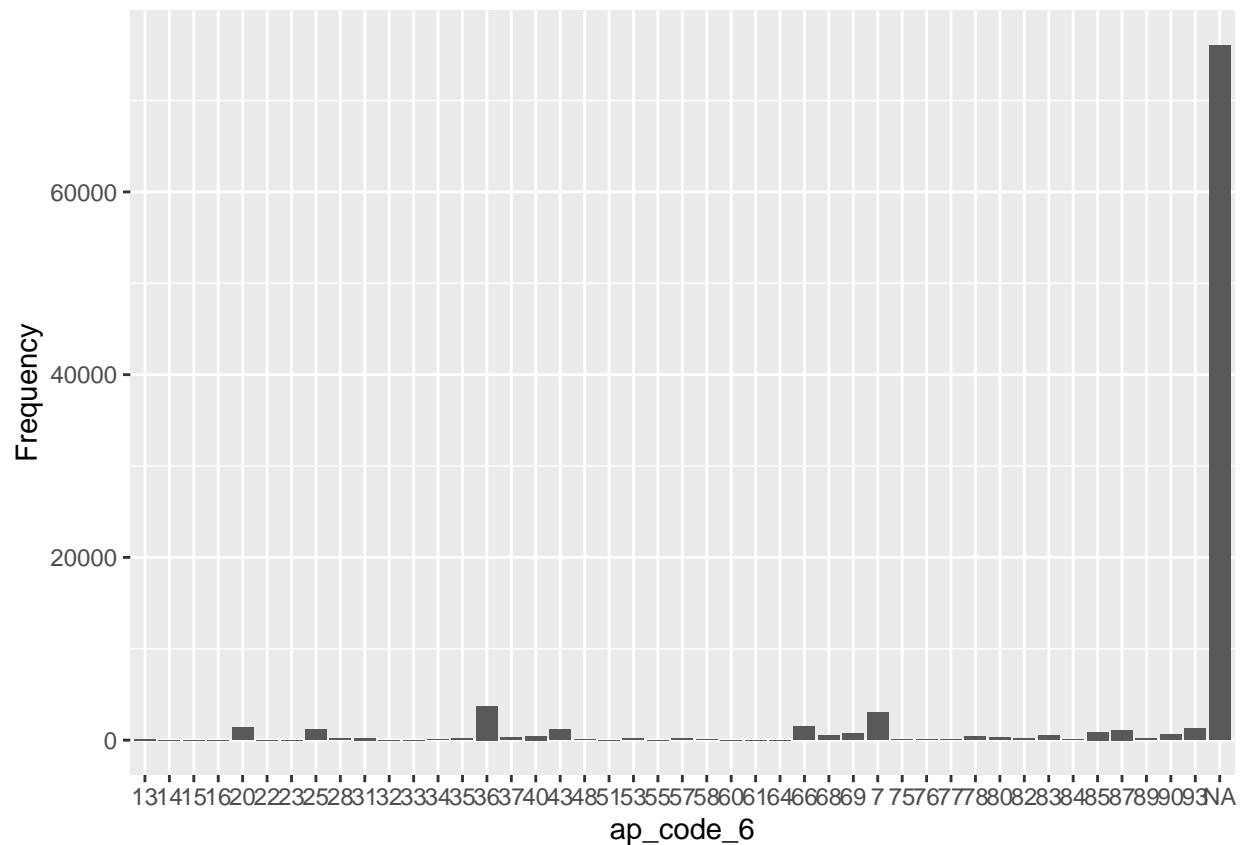


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_5, type: numeric, requested: yes
[1] Values (11 unique): 0, NA, 2, 4, 5, ...
[1] Missing: 33.2%
  Group.1 ap_essay_flag_5
1    2006    0.9990588235
2    2007    0.9985261606
3    2008    0.9955391478
4    2009    0.9974827991
5    2010    0.9970638232
6    2011    0.0004249292
7    2012    0.0005745475
8    2013    0.0003934942
9    2014    0.0037760417
10   2015    0.0037821483
11   2016    0.0055980192
12   2017    0.0028815478
13   2018    0.0113871636
```

Warning: Removed 32260 rows containing non-finite values ('stat_count()').

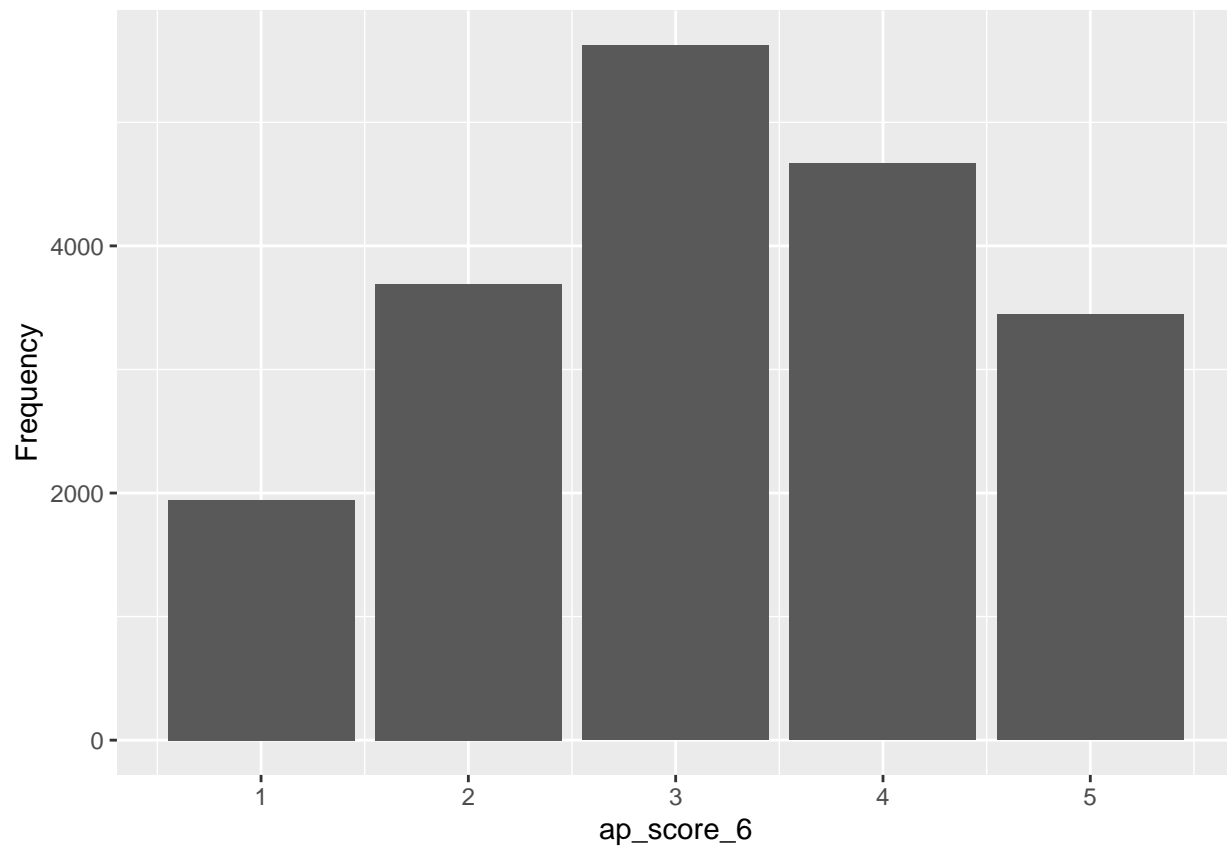


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_6, type: character, requested: yes
[1] Values (45 unique): NA, 36, 69, 25, 20, ...
[1] Missing: 78.4%
[1] Most missing: 2007 100%, Least missing: 2017 65.5%
```



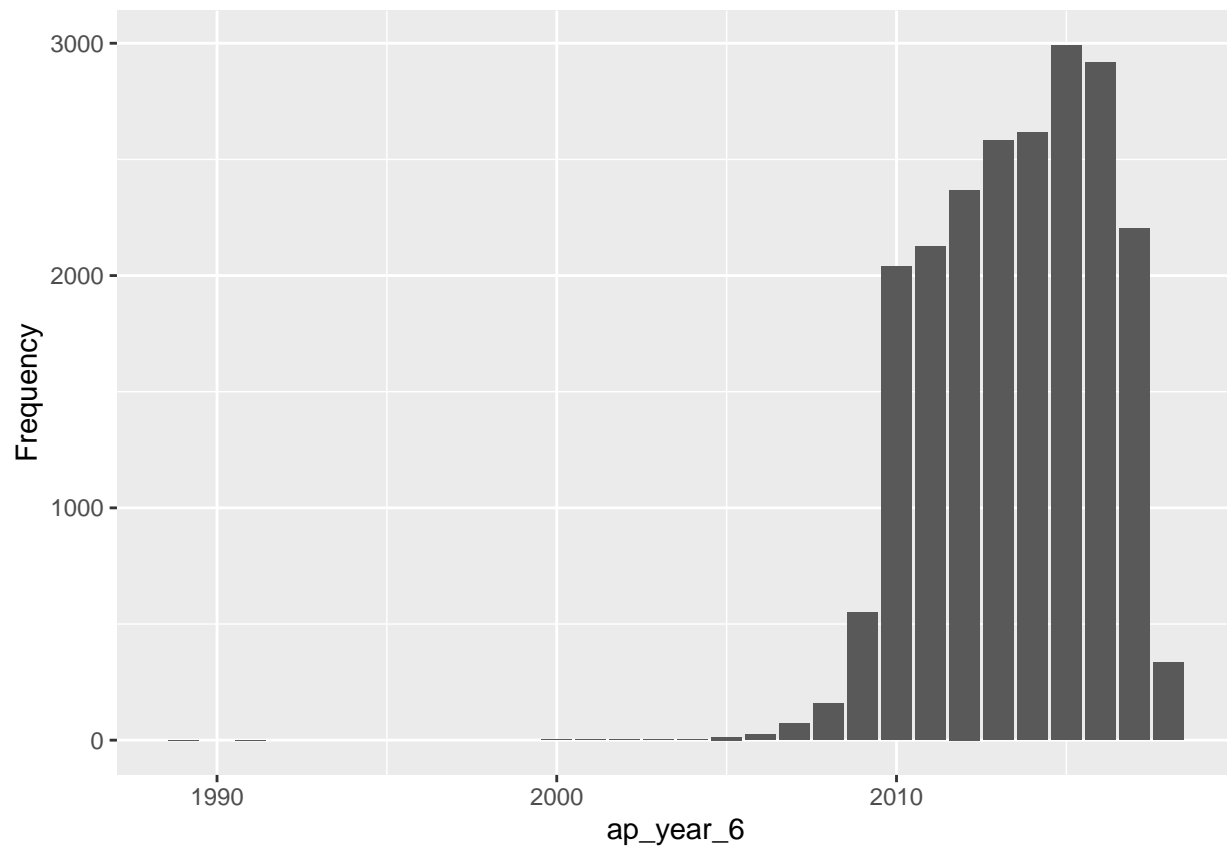
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_6, type: numeric, requested: yes
[1] Values (6 unique): NA, 4, 5, 3, 2, ...
[1] Missing: 80%
[1] Most missing: 2007 100%, Least missing: 2017 66.7%
```

Warning: Removed 77682 rows containing non-finite values ('stat_count()').



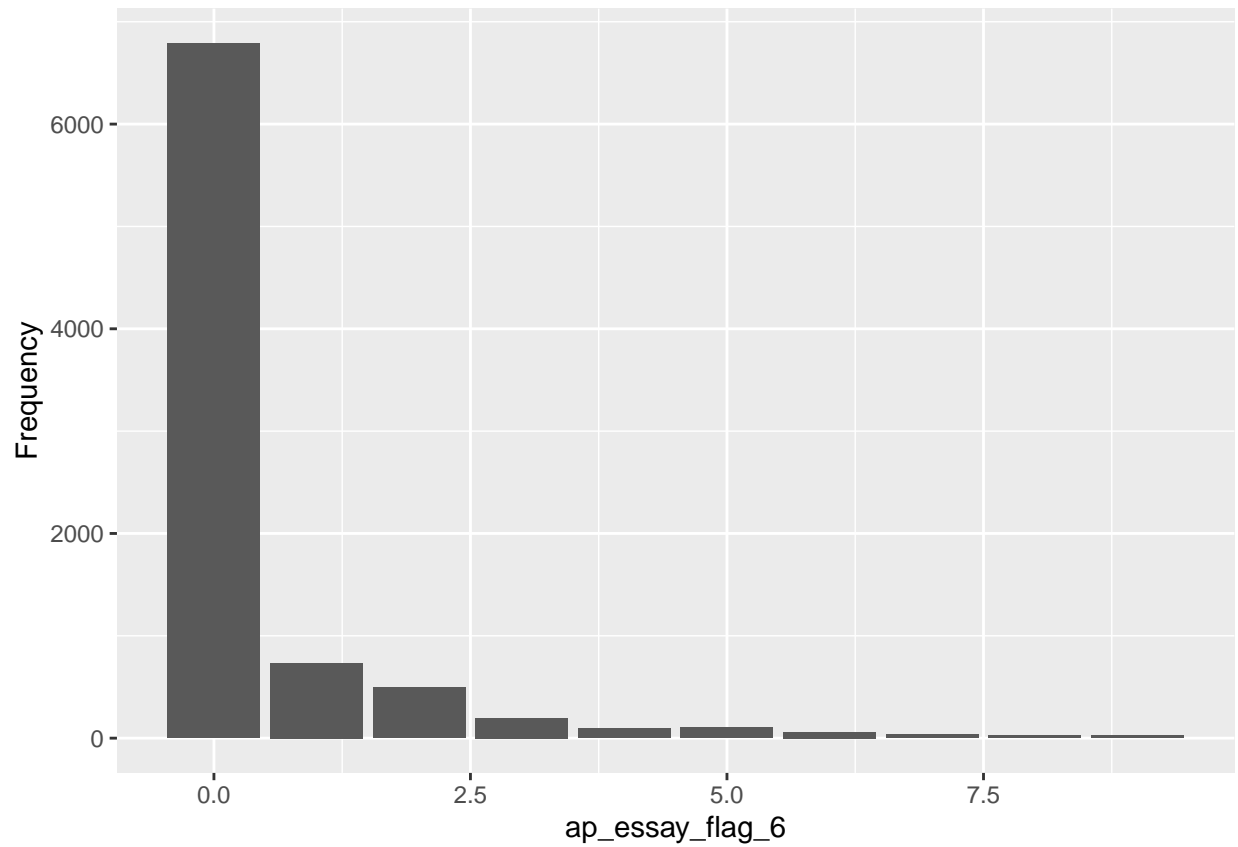
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_6, type: integer, requested: yes
[1] Values (22 unique): NA, 2017, 2015, 2018, 2016, ...
[1] Missing: 78.4%
[1] Most missing: 2007 100%, Least missing: 2017 65.5%
```

Warning: Removed 76040 rows containing non-finite values ('stat_count()').

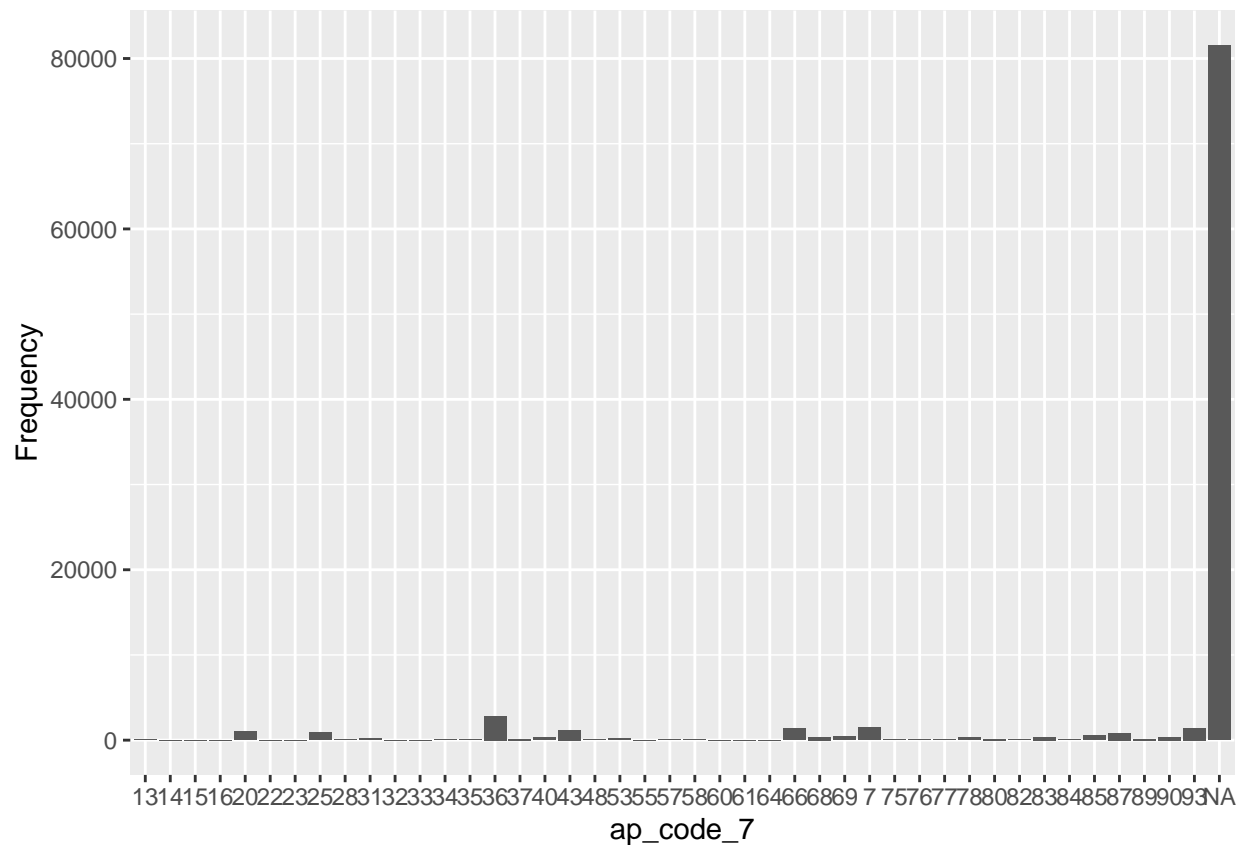


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_6, type: numeric, requested: yes
[1] Values (11 unique): NA, 0, 3, 2, 1, ...
[1] Missing: 91.2%
[1] Most missing: 2018 100%, Least missing: 2014 68.8%
```

Warning: Removed 88489 rows containing non-finite values ('stat_count()').

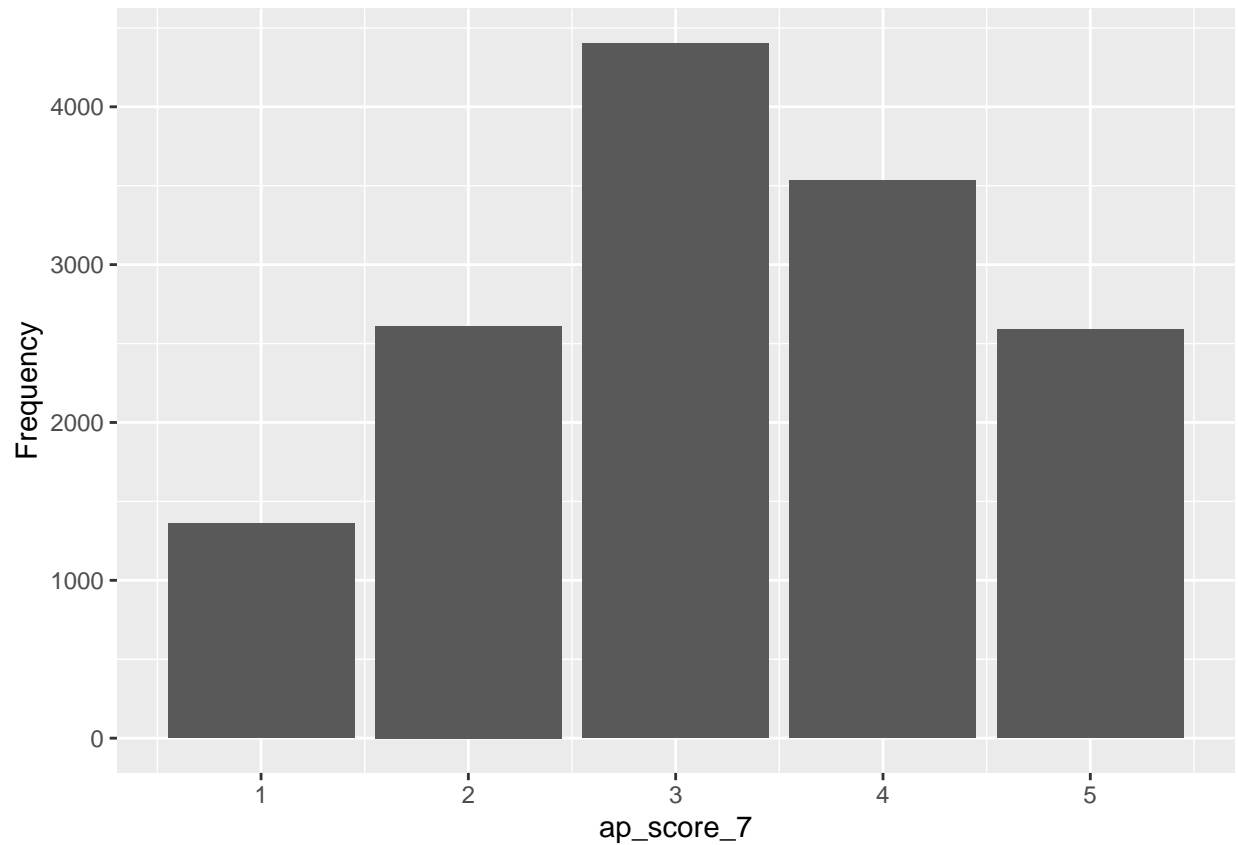


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_7, type: character, requested: yes
[1] Values (44 unique): NA, 66, 36, 7, 25, ...
[1] Missing: 84%
[1] Most missing: 2007 100%, Least missing: 2017 73.9%
```



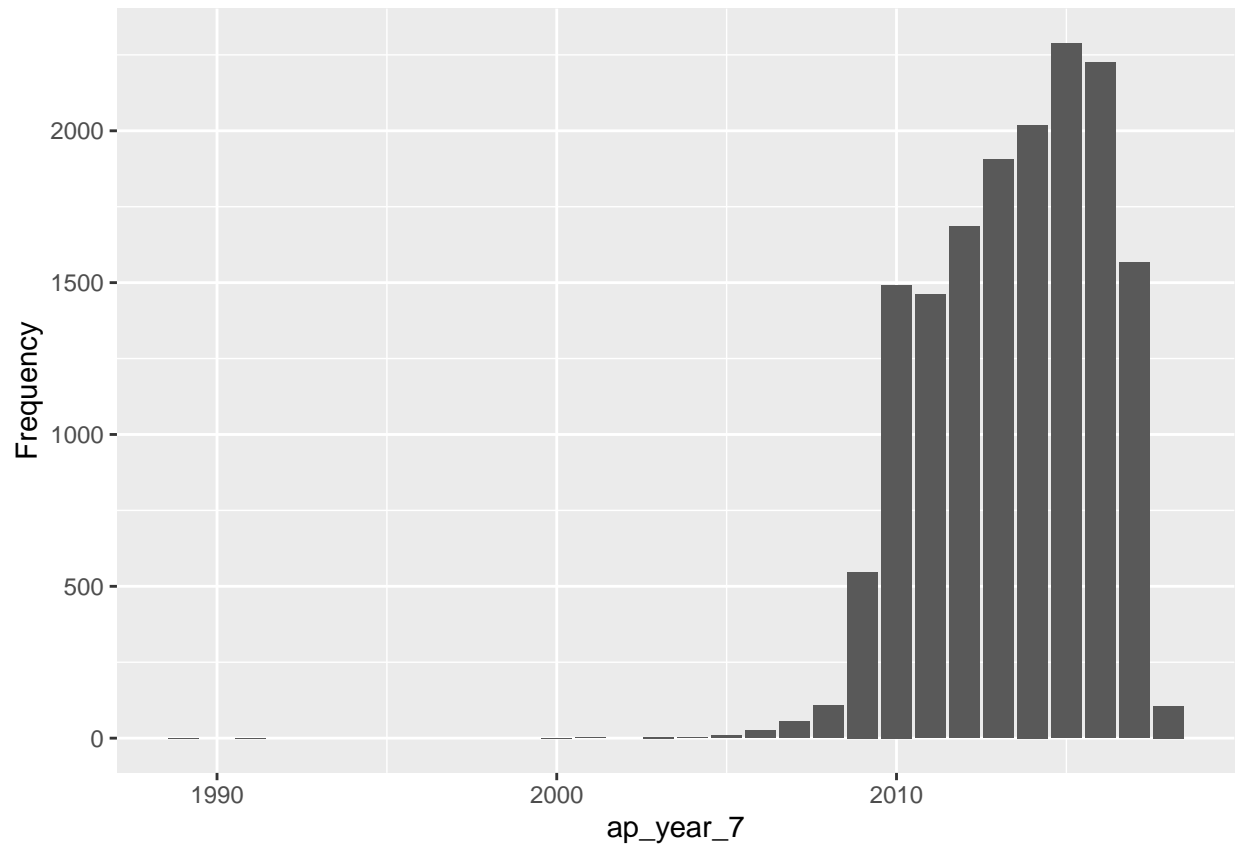
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_7, type: numeric, requested: yes
[1] Values (6 unique): NA, 3, 5, 4, 2, ...
[1] Missing: 85.1%
[1] Most missing: 2007 100%, Least missing: 2017 74.5%
```

Warning: Removed 82548 rows containing non-finite values ('stat_count()').



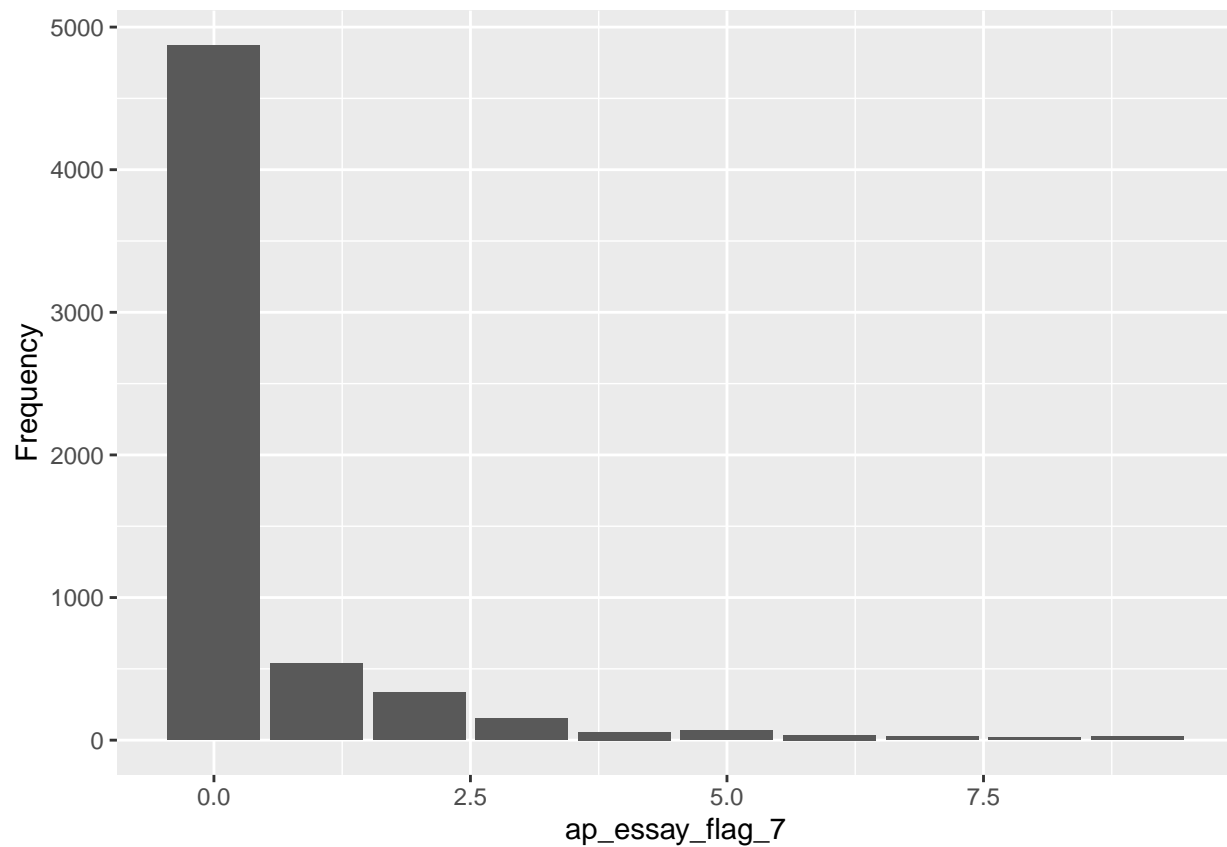
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_7, type: integer, requested: yes
[1] Values (21 unique): NA, 2017, 2016, 2015, 2014, ...
[1] Missing: 84%
[1] Most missing: 2007 100%, Least missing: 2017 73.9%
```

Warning: Removed 81545 rows containing non-finite values ('stat_count()').

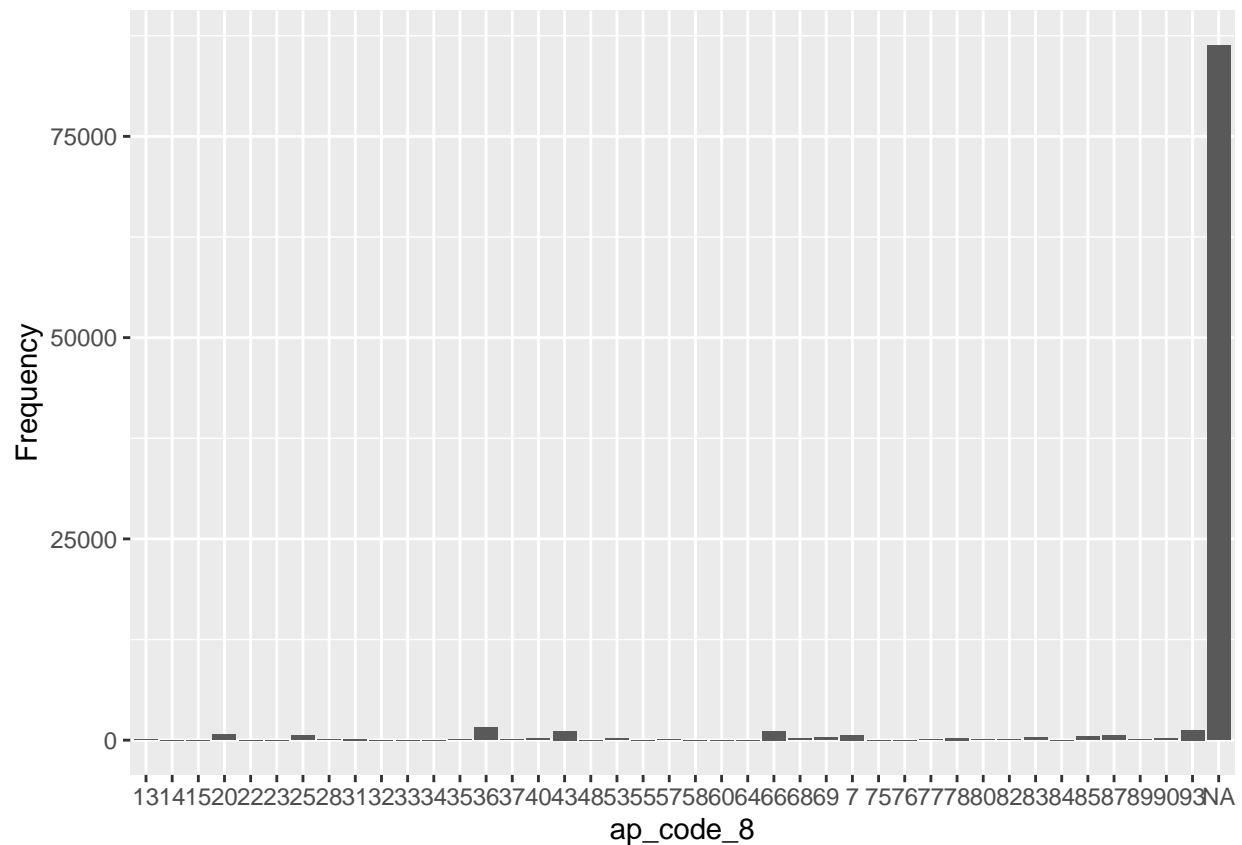


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_7, type: numeric, requested: yes
[1] Values (11 unique): NA, 0, 1, 4, 3, ...
[1] Missing: 93.7%
[1] Most missing: 2018 100%, Least missing: 2014 77%
```

Warning: Removed 90916 rows containing non-finite values ('stat_count()').

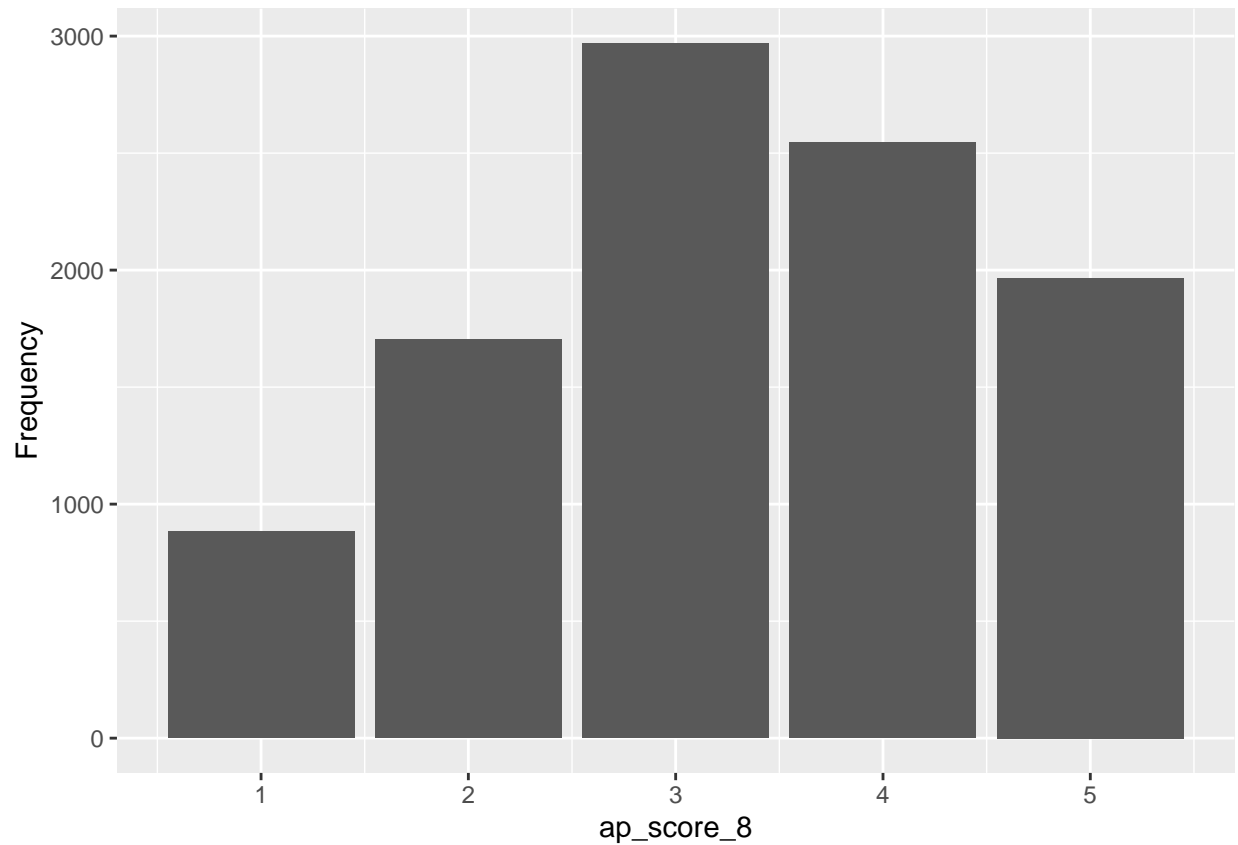


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_8, type: character, requested: yes
[1] Values (42 unique): NA, 93, 43, 66, 87, ...
[1] Missing: 88.9%
[1] Most missing: 2006 100%, Least missing: 2018 80.8%
```



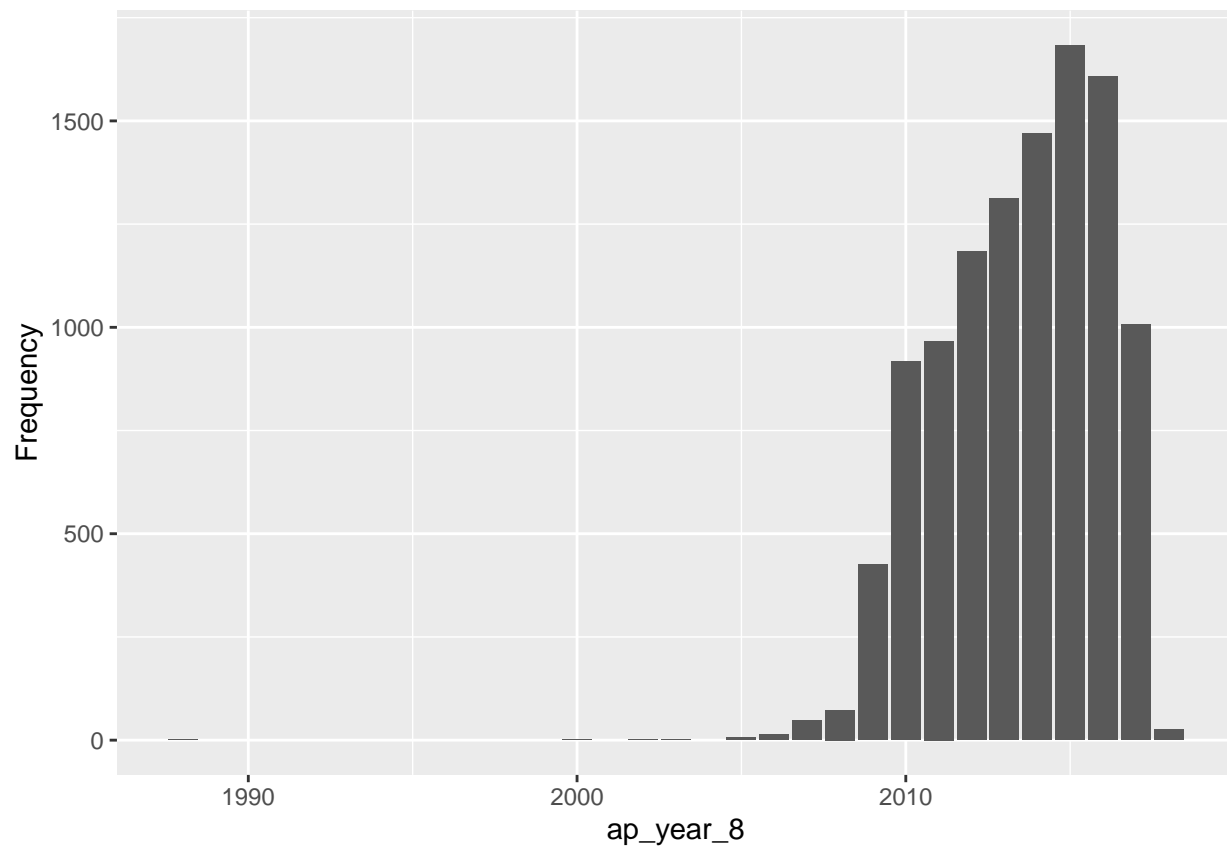
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_8, type: numeric, requested: yes
[1] Values (6 unique): NA, 4, 3, 2, 5, ...
[1] Missing: 89.6%
[1] Most missing: 2006 100%, Least missing: 2018 81%
```

Warning: Removed 86983 rows containing non-finite values ('stat_count()').



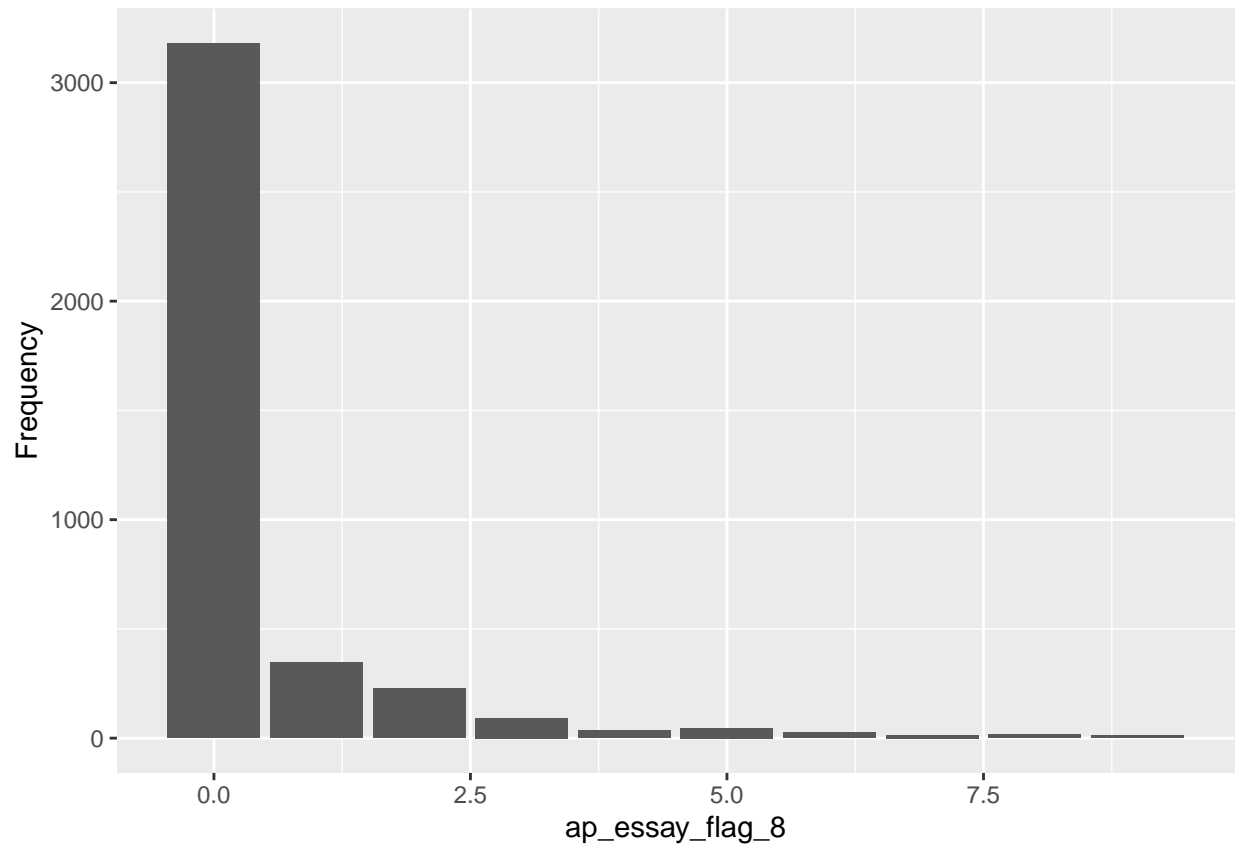
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_8, type: integer, requested: yes
[1] Values (19 unique): NA, 2016, 2017, 2015, 2014, ...
[1] Missing: 88.9%
[1] Most missing: 2006 100%, Least missing: 2018 80.8%
```

Warning: Removed 86306 rows containing non-finite values ('stat_count()').

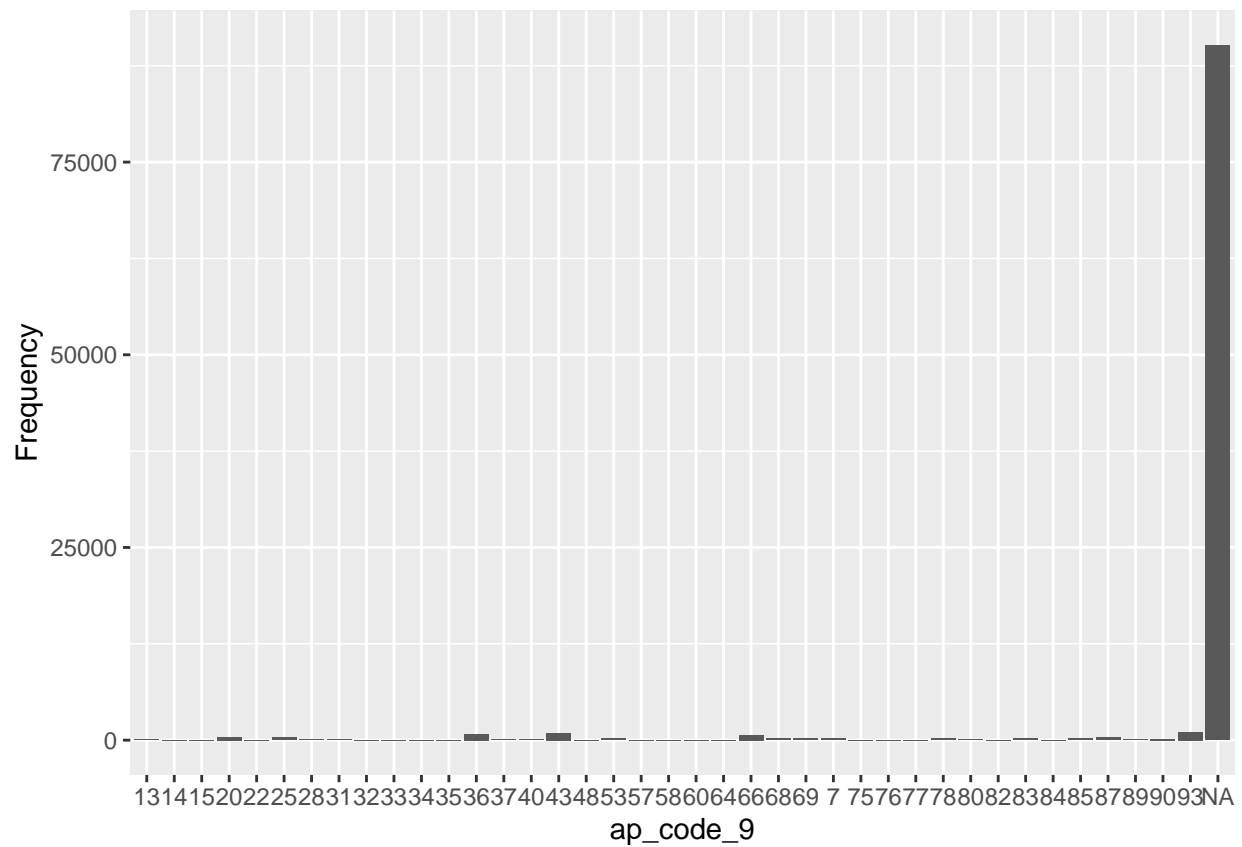


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_8, type: numeric, requested: yes
[1] Values (11 unique): NA, 0, 3, 1, 5, ...
[1] Missing: 95.9%
[1] Most missing: 2006 100%, Least missing: 2014 84.5%
```

Warning: Removed 93056 rows containing non-finite values ('stat_count()').

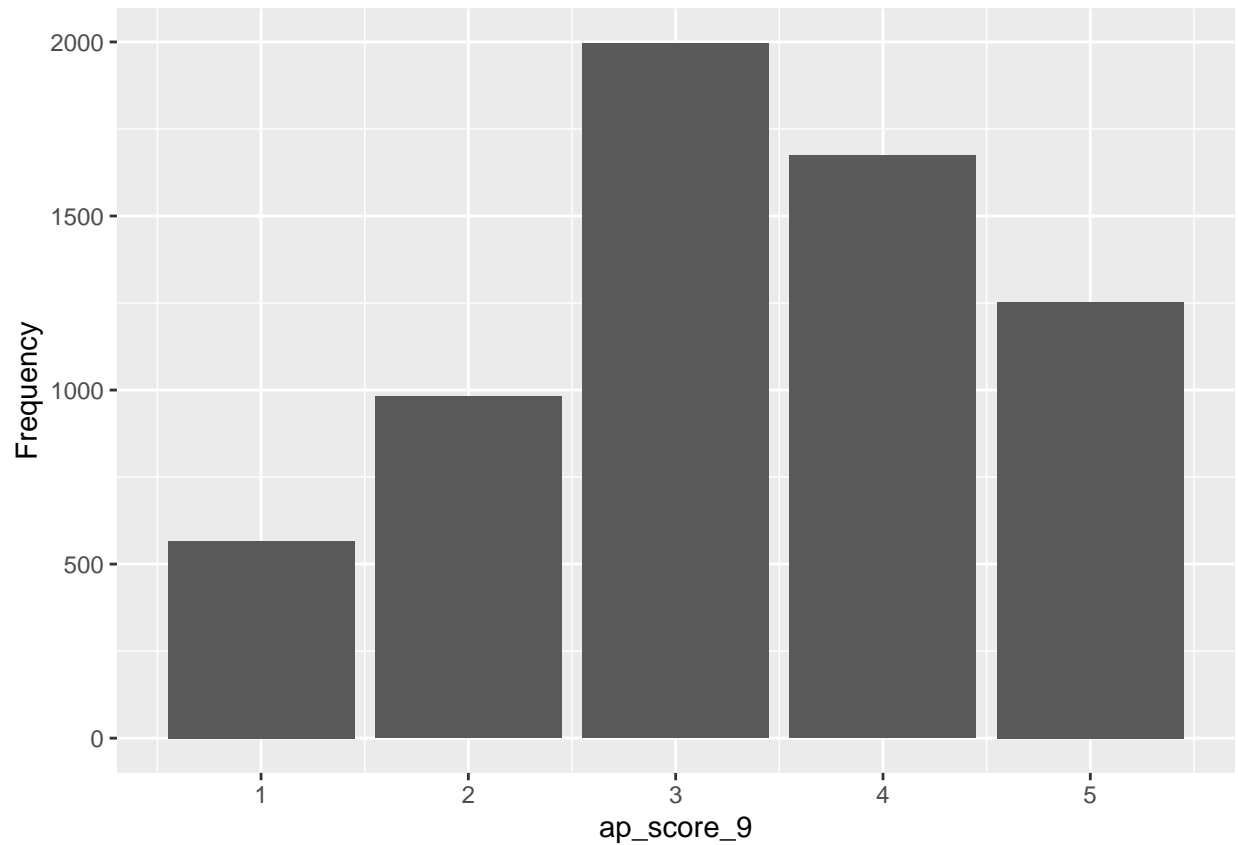


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_9, type: character, requested: yes
[1] Values (40 unique): NA, 53, 43, 93, 36, ...
[1] Missing: 92.9%
[1] Most missing: 2006 100%, Least missing: 2018 86.9%
```



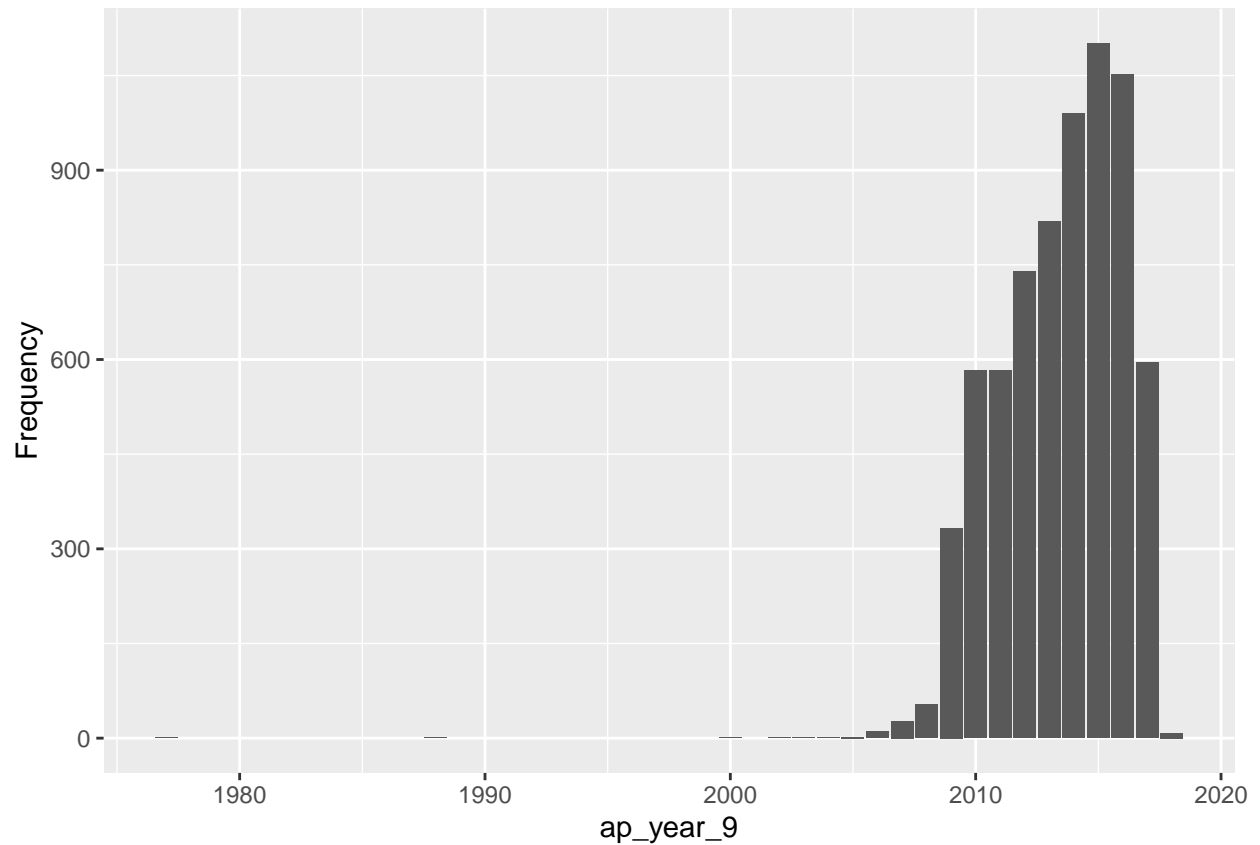
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_9, type: numeric, requested: yes
[1] Values (6 unique): NA, 3, 2, 4, 1, ...
[1] Missing: 93.3%
[1] Most missing: 2006 100%, Least missing: 2018 87.1%
```

Warning: Removed 90579 rows containing non-finite values ('stat_count()').



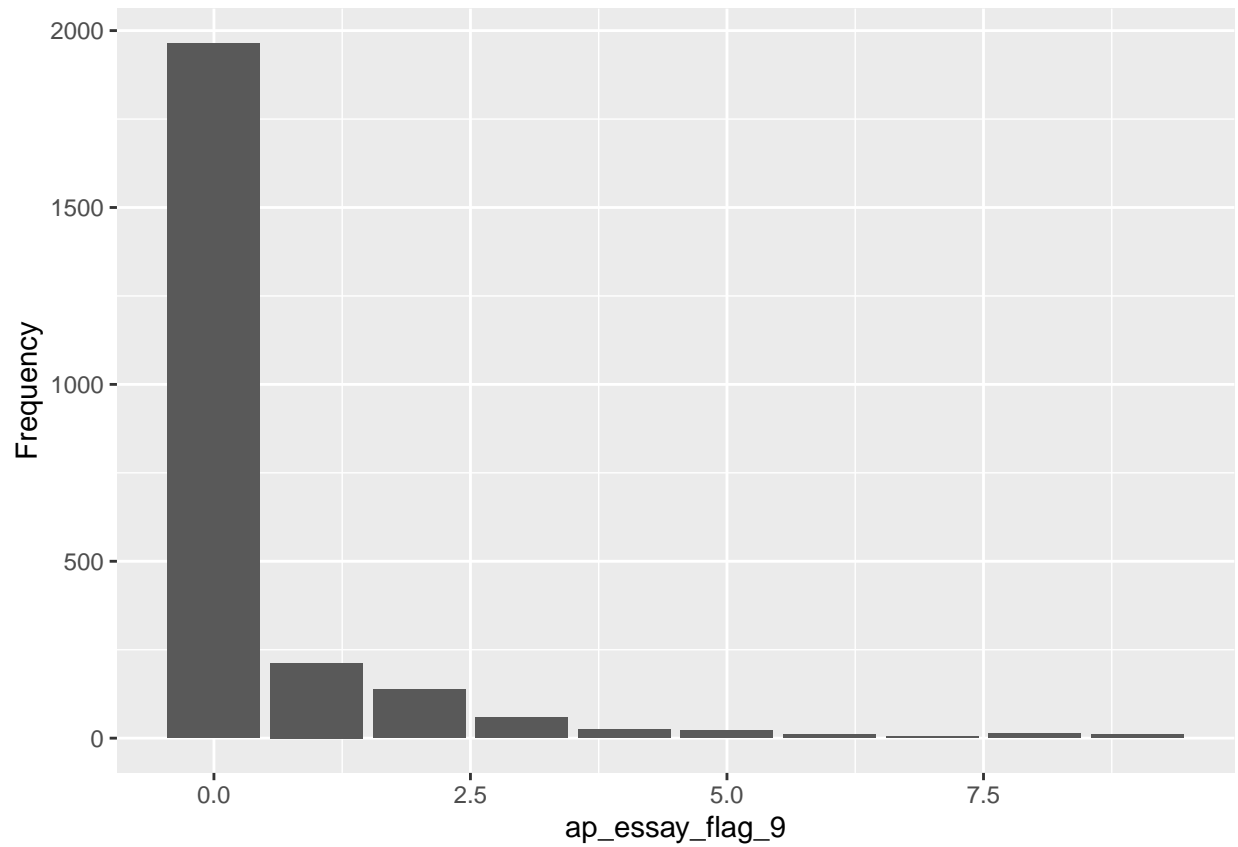
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_9, type: integer, requested: yes
[1] Values (21 unique): NA, 2015, 2016, 2014, 2017, ...
[1] Missing: 92.9%
[1] Most missing: 2006 100%, Least missing: 2018 86.9%
```

Warning: Removed 90148 rows containing non-finite values ('stat_count()').

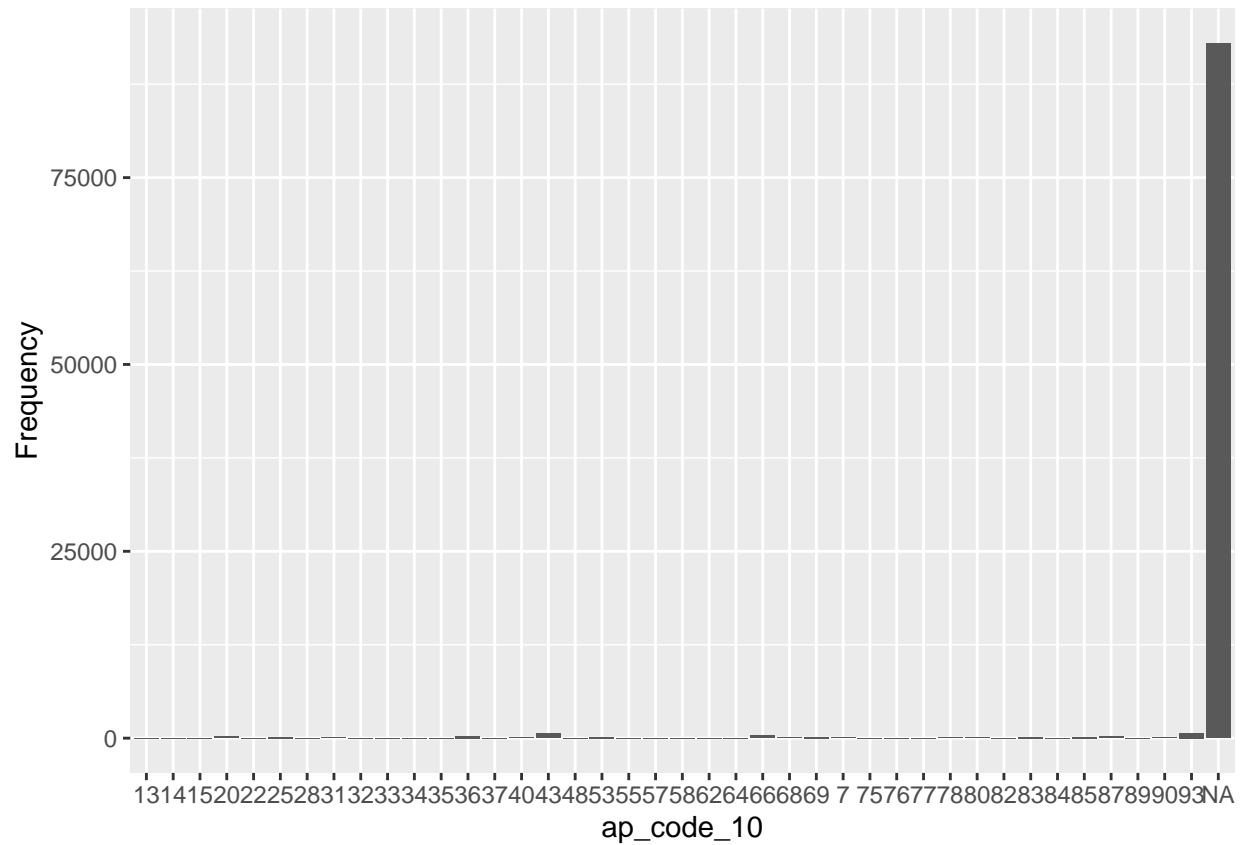


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_9, type: numeric, requested: yes
[1] Values (11 unique): NA, 0, 3, 1, 2, ...
[1] Missing: 97.5%
[1] Most missing: 2006 100%, Least missing: 2014 90.1%
```

Warning: Removed 94591 rows containing non-finite values ('stat_count()').

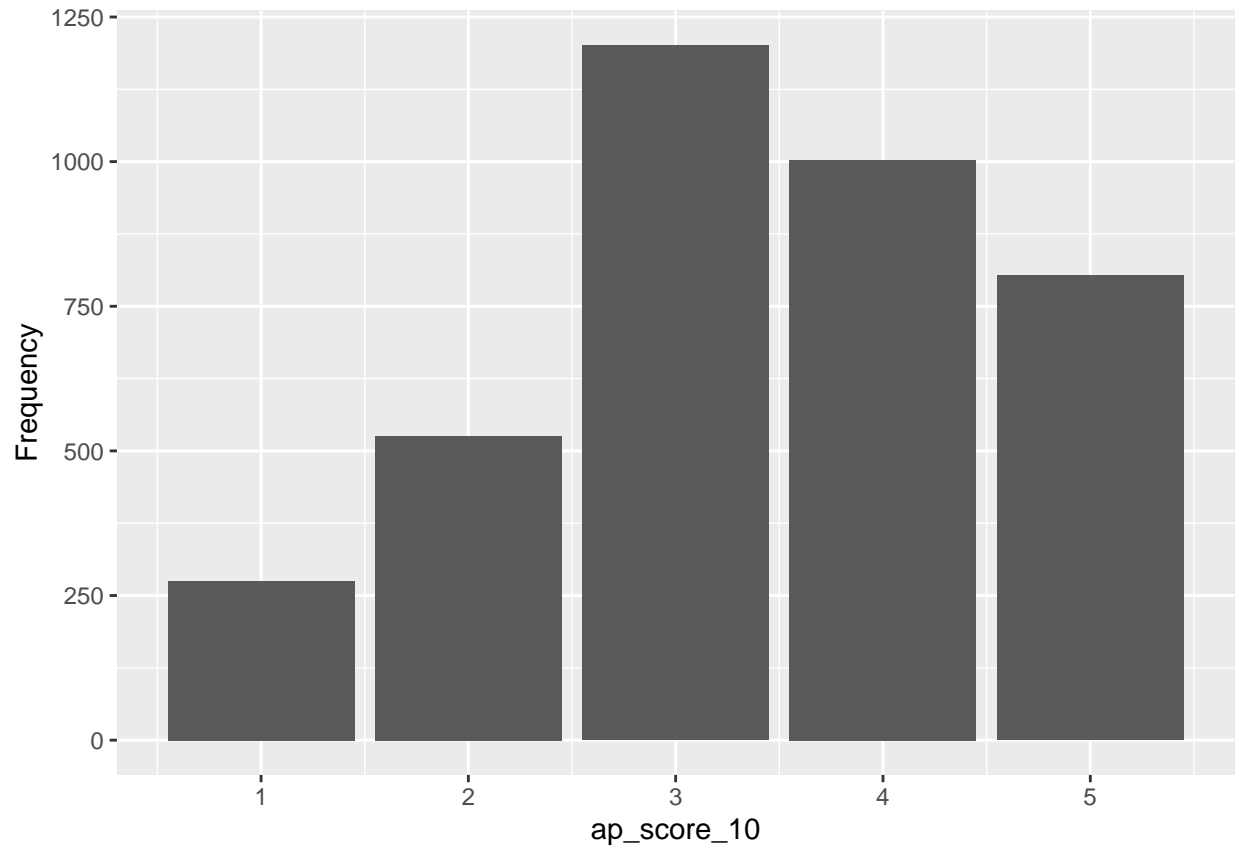


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_10, type: character, requested: yes
[1] Values (41 unique): NA, 87, 83, 25, 40, ...
[1] Missing: 95.8%
[1] Most missing: 2006 100%, Least missing: 2018 91.6%
```



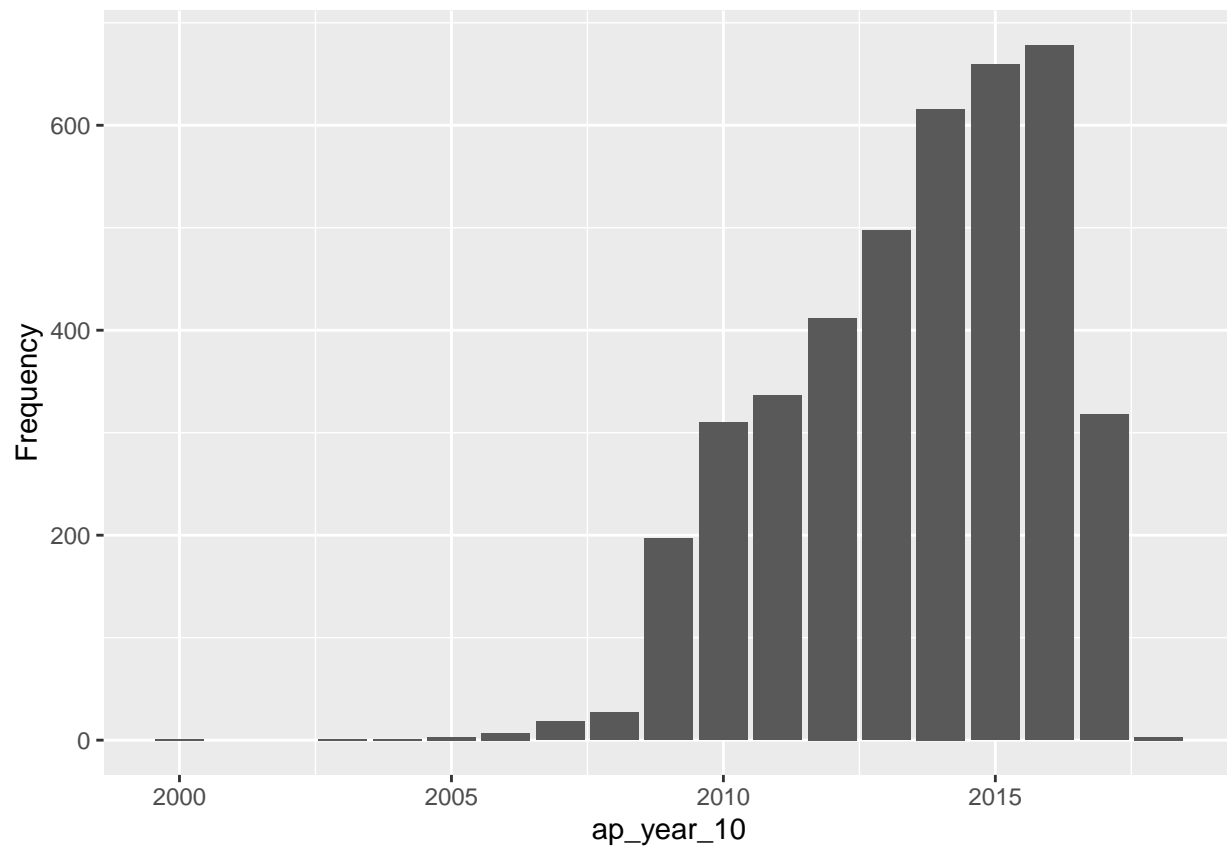
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_10, type: numeric, requested: yes
[1] Values (6 unique): NA, 4, 3, 5, 1, ...
[1] Missing: 96.1%
[1] Most missing: 2006 100%, Least missing: 2018 91.6%
```

Warning: Removed 93241 rows containing non-finite values ('stat_count()').



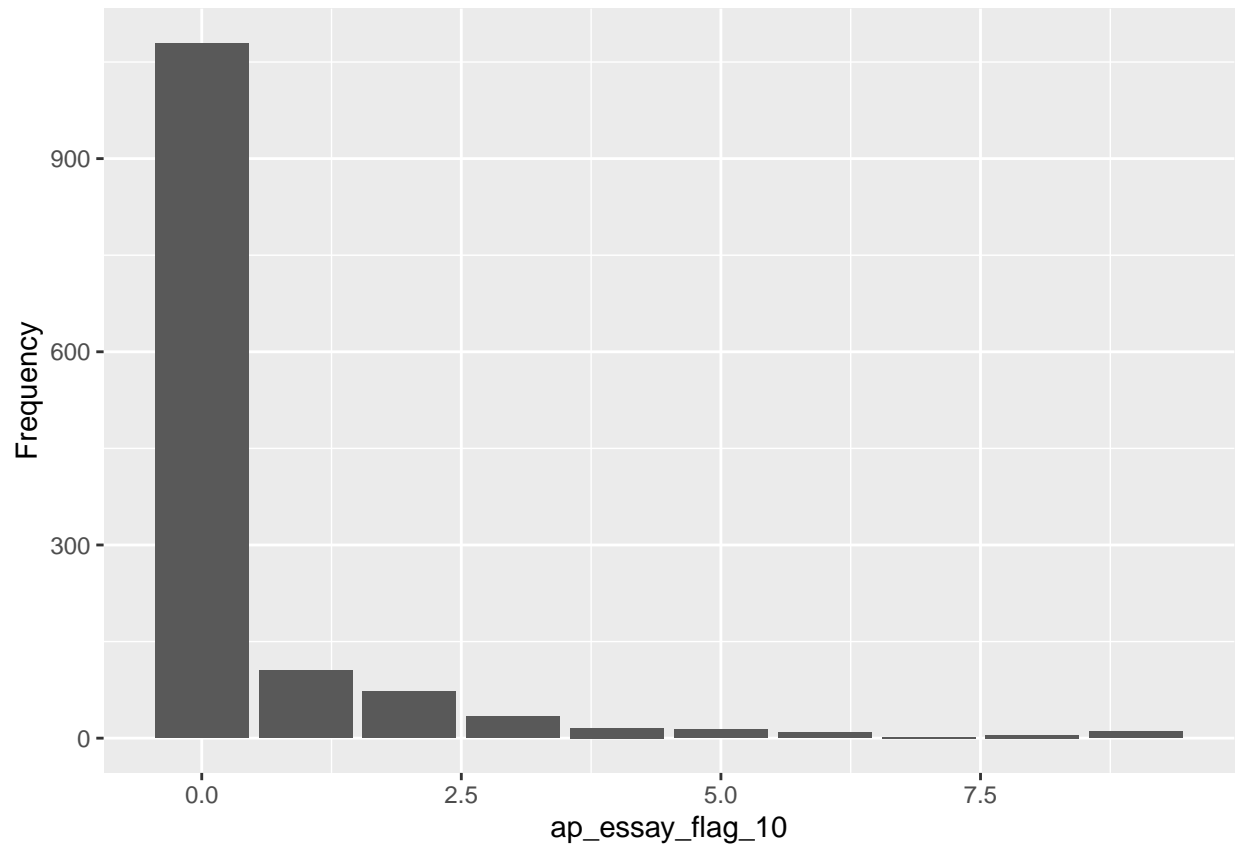
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_10, type: integer, requested: yes
[1] Values (18 unique): NA, 2016, 2017, 2015, 2013, ...
[1] Missing: 95.8%
[1] Most missing: 2006 100%, Least missing: 2018 91.6%
```

Warning: Removed 92965 rows containing non-finite values ('stat_count()').

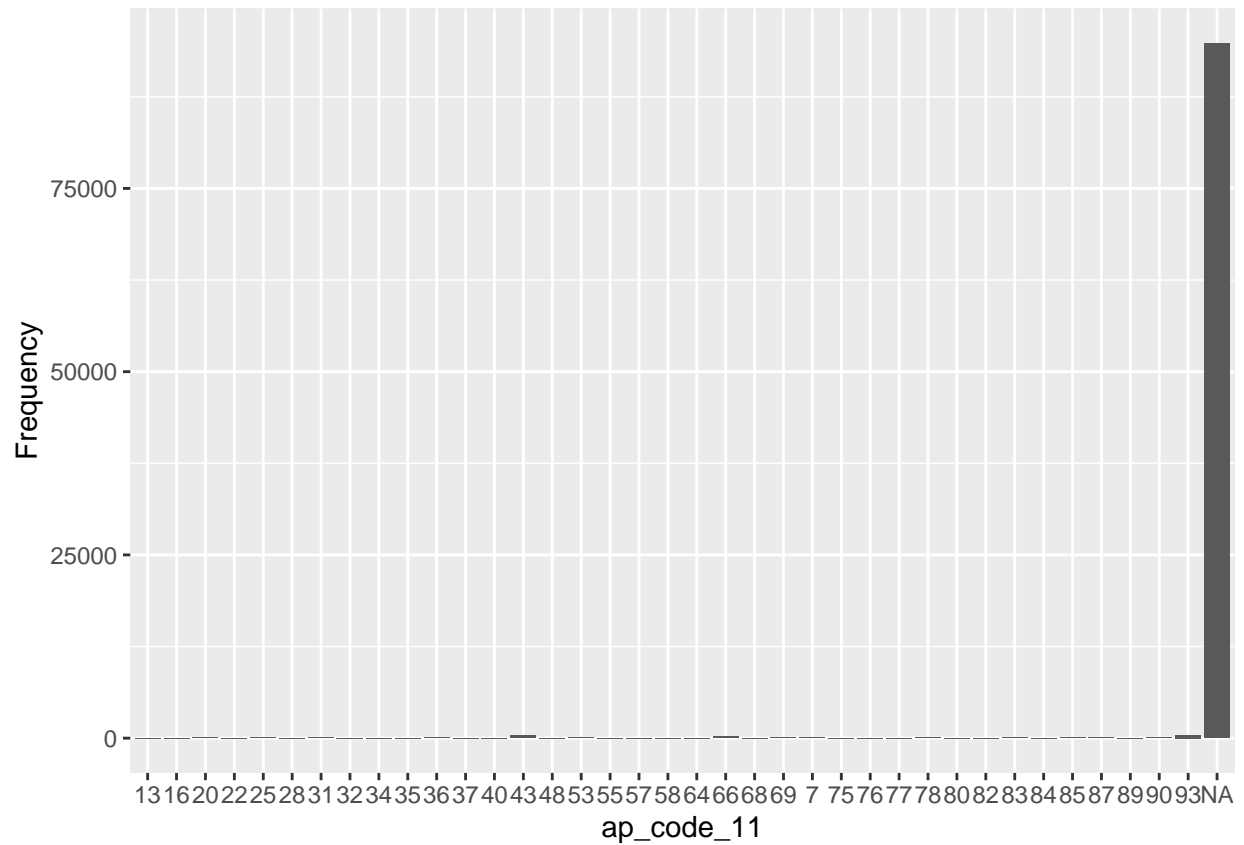


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_10, type: numeric, requested: yes
[1] Values (11 unique): NA, 0, 1, 2, 3, ...
[1] Missing: 98.6%
[1] Most missing: 2006 100%, Least missing: 2014 94.5%
```

Warning: Removed 95704 rows containing non-finite values ('stat_count()').

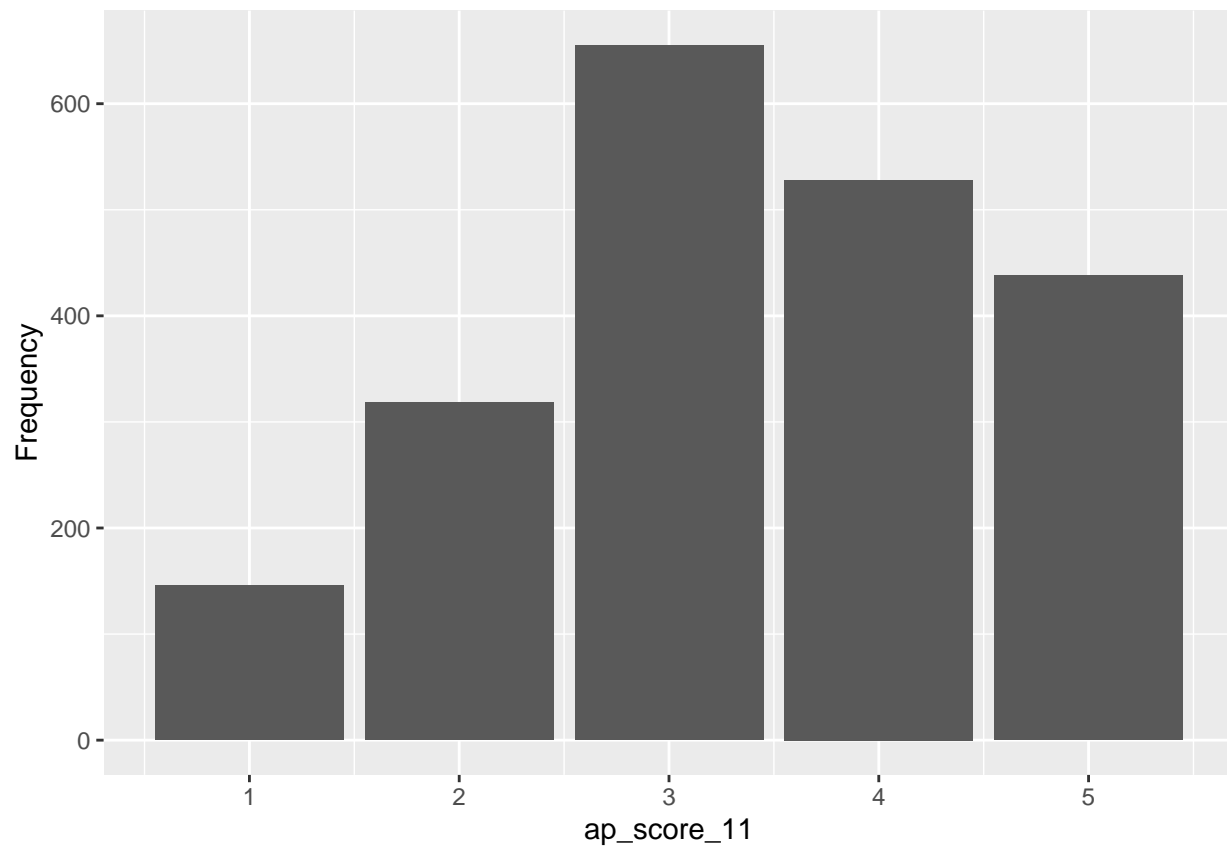


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_11, type: character, requested: yes
[1] Values (38 unique): NA, 53, 43, 93, 83, ...
[1] Missing: 97.7%
[1] Most missing: 2006 100%, Least missing: 2018 94.7%
```



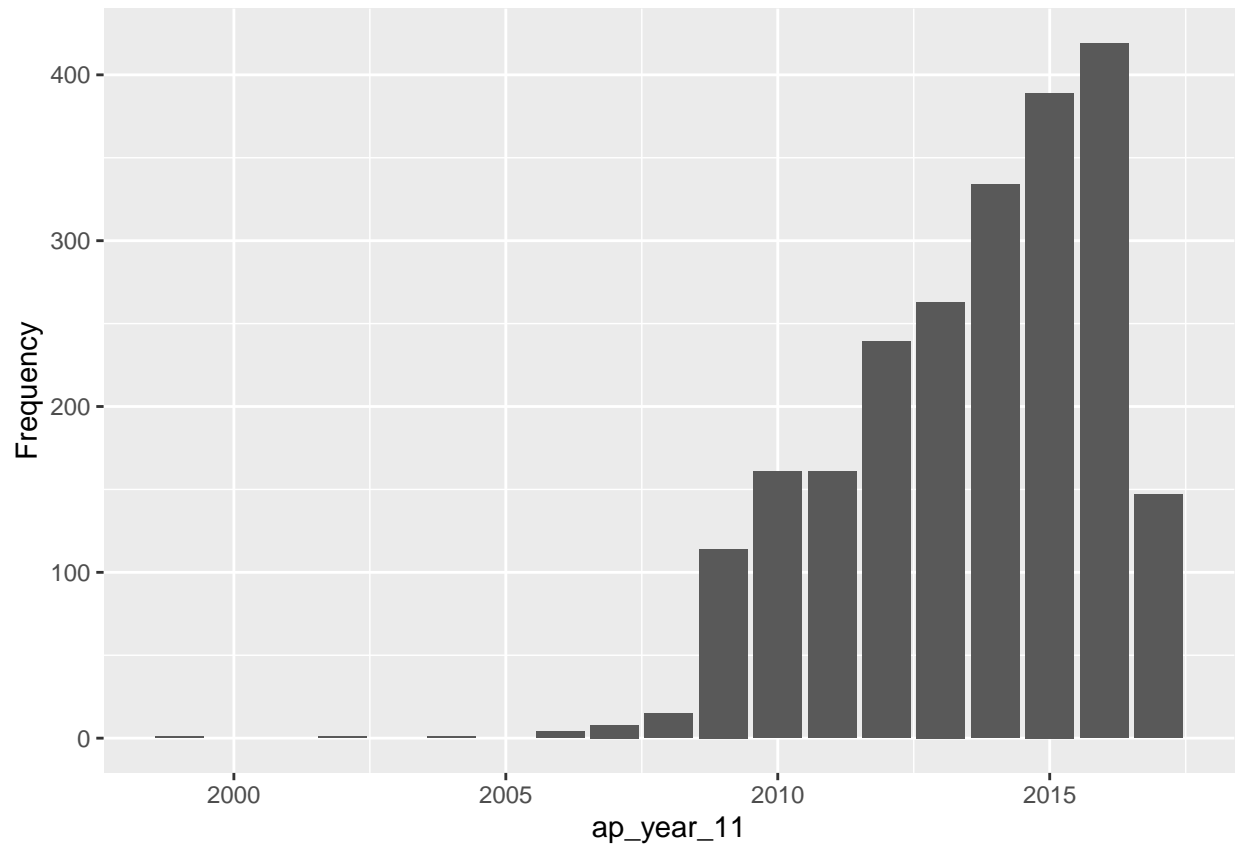
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_11, type: numeric, requested: yes
[1] Values (6 unique): NA, 1, 4, 3, 2, ...
[1] Missing: 97.9%
[1] Most missing: 2006 100%, Least missing: 2018 94.8%
```

Warning: Removed 94964 rows containing non-finite values ('stat_count()').



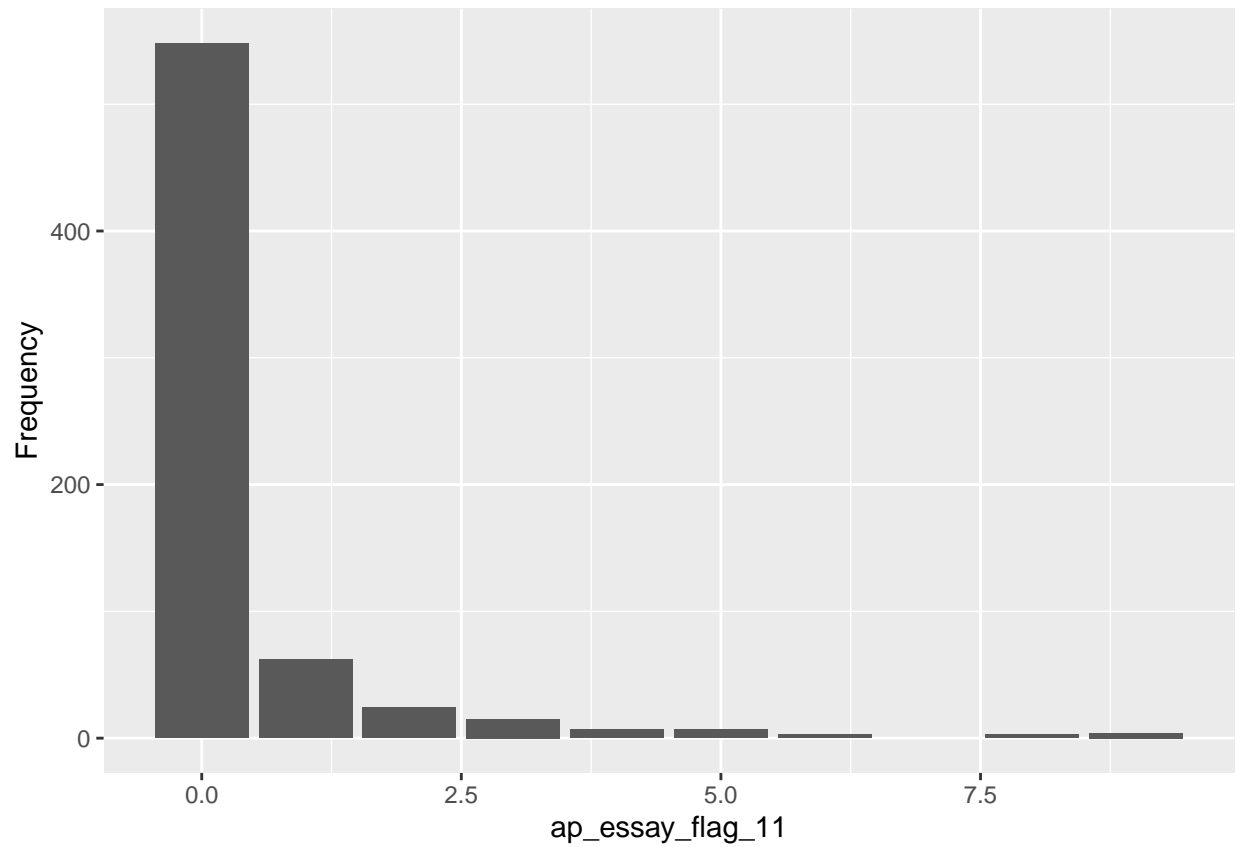
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_11, type: integer, requested: yes
[1] Values (16 unique): NA, 2015, 2016, 2017, 2014, ...
[1] Missing: 97.7%
[1] Most missing: 2006 100%, Least missing: 2018 94.7%
```

Warning: Removed 94792 rows containing non-finite values ('stat_count()').

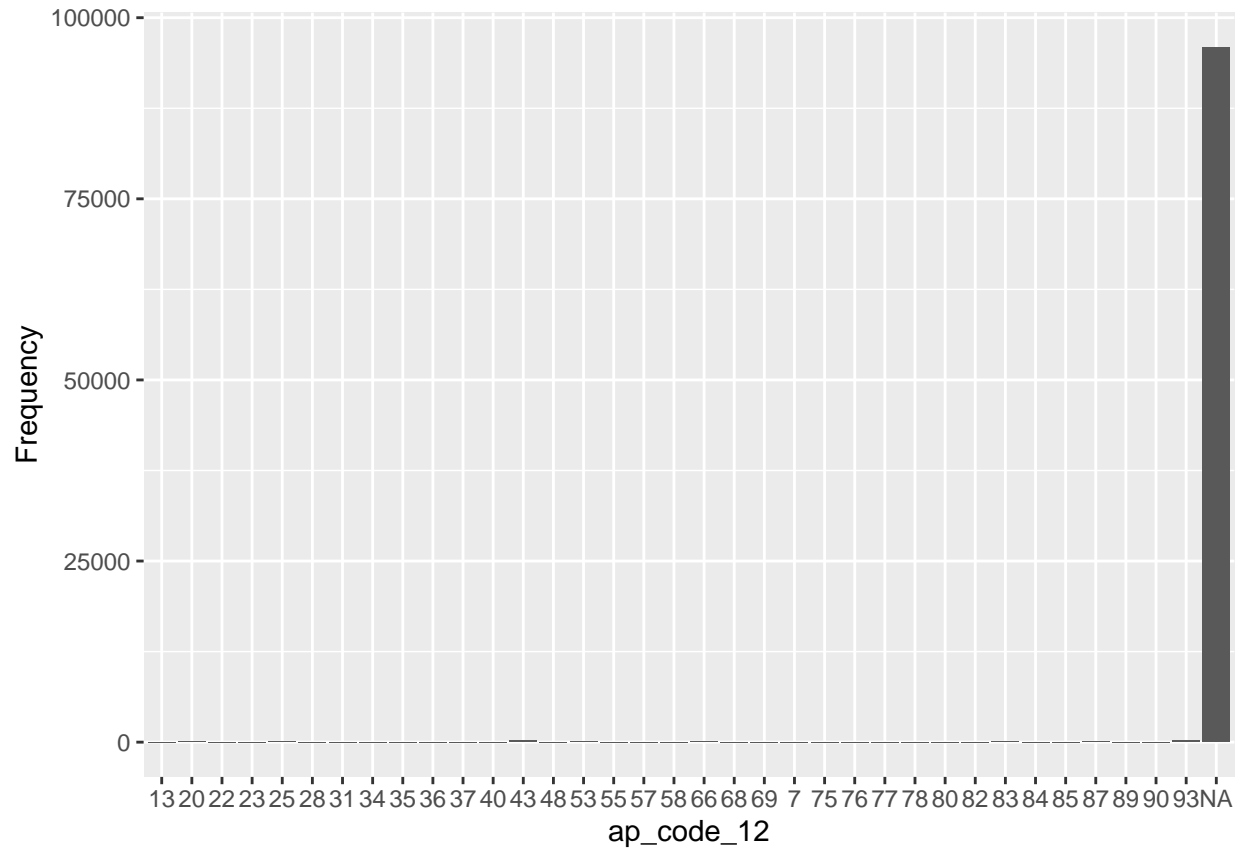


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_11, type: numeric, requested: yes
[1] Values (10 unique): NA, 0, 1, 2, 5, ...
[1] Missing: 99.3%
```

Warning: Removed 96376 rows containing non-finite values ('stat_count()').

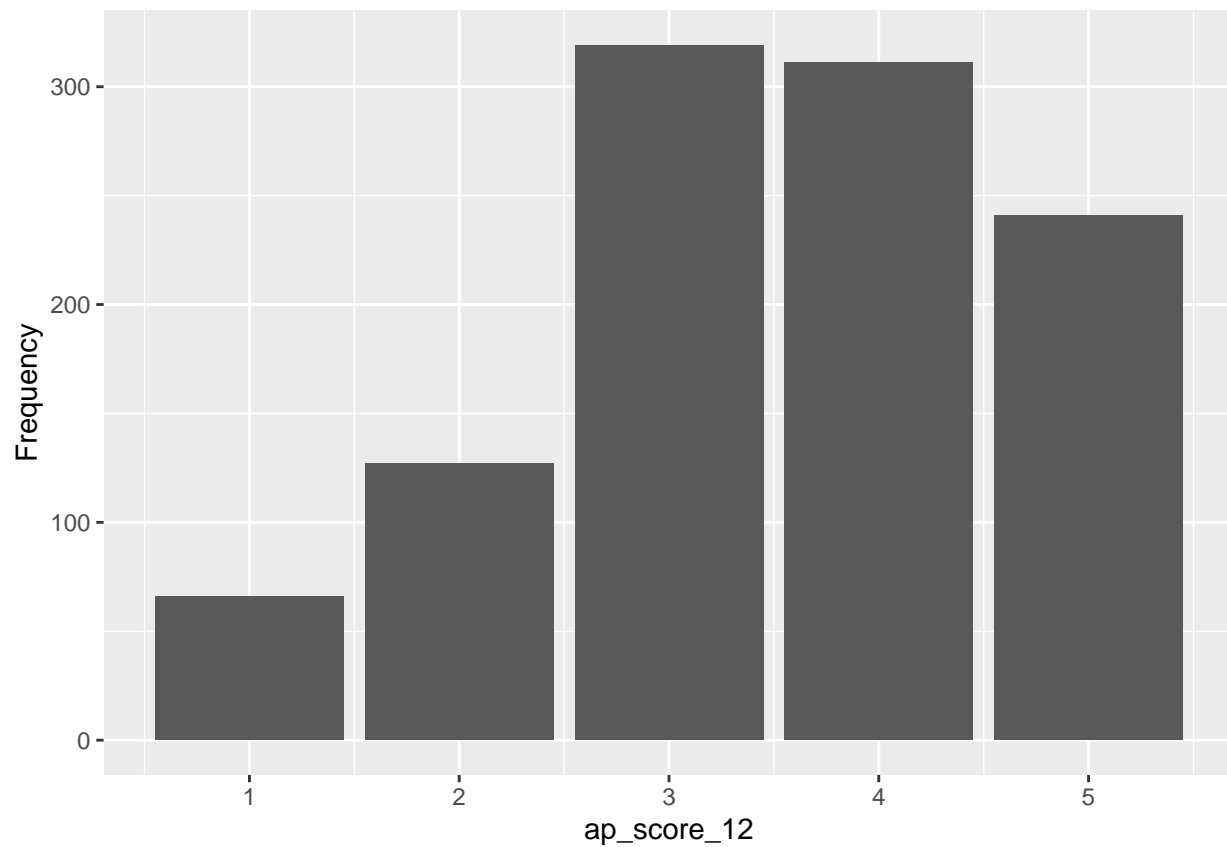


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_12, type: character, requested: yes
[1] Values (36 unique): NA, 53, 85, 43, 93, ...
[1] Missing: 98.8%
[1] Most missing: 2006 100%, Least missing: 2018 97%
```



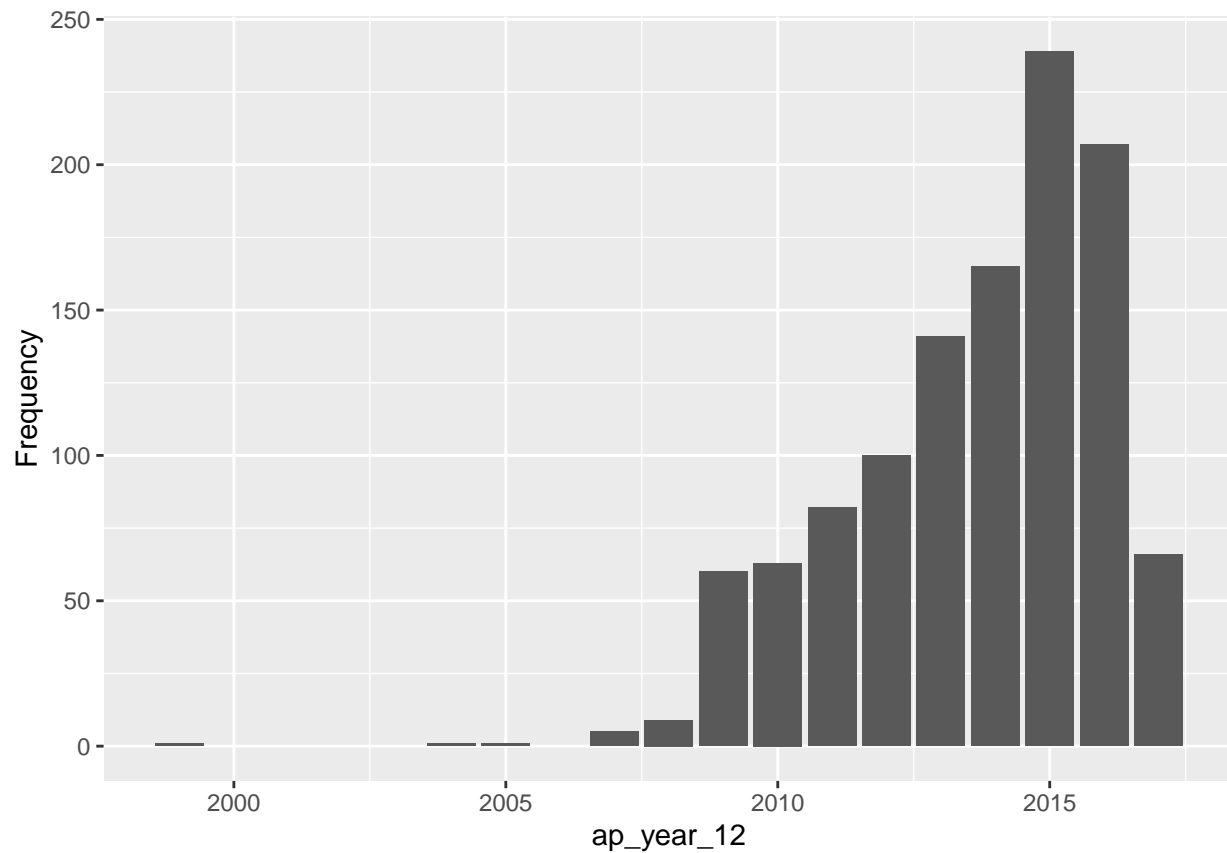
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_12, type: numeric, requested: yes
[1] Values (6 unique): NA, 3, 4, 5, 2, ...
[1] Missing: 98.9%
[1] Most missing: 2006 100%, Least missing: 2018 97.1%
```

Warning: Removed 95985 rows containing non-finite values ('stat_count()').



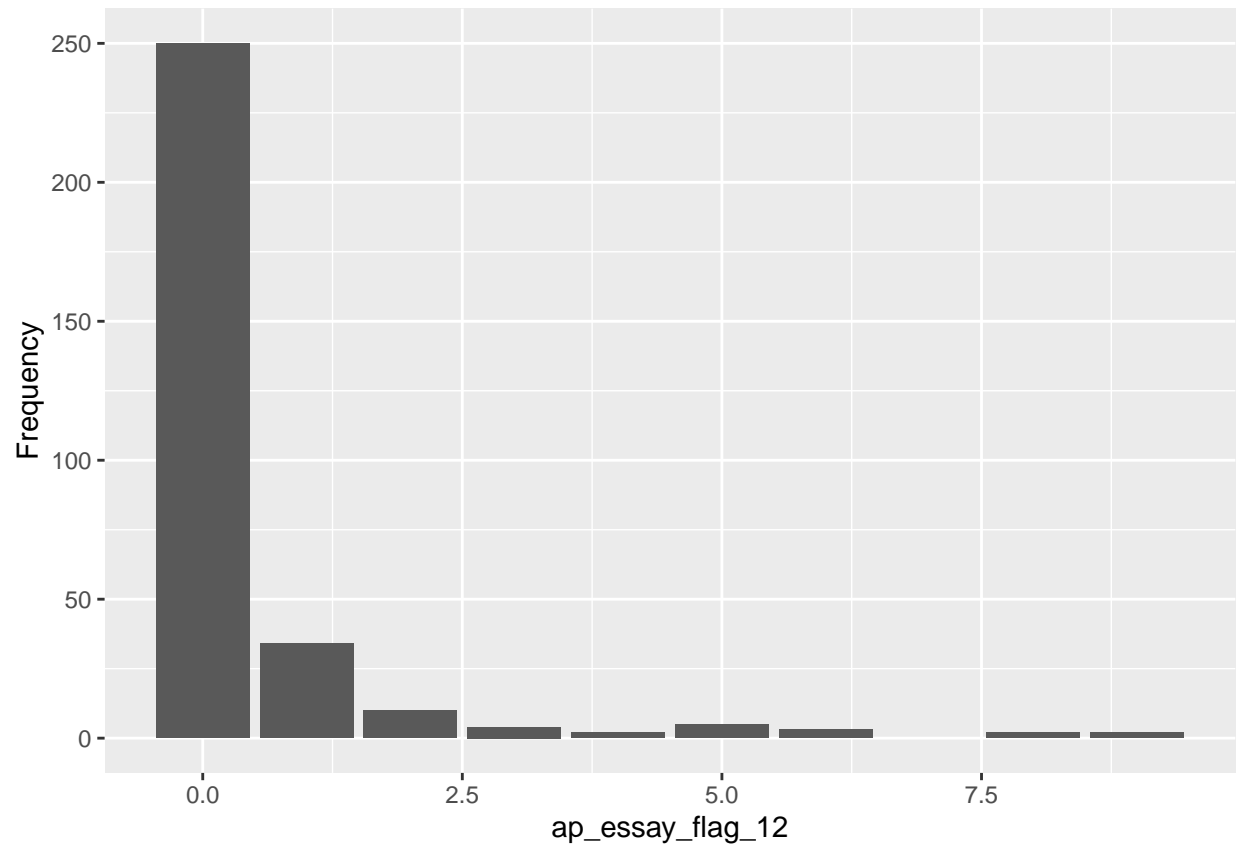
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_12, type: integer, requested: yes
[1] Values (15 unique): NA, 2015, 2016, 2014, 2017, ...
[1] Missing: 98.8%
[1] Most missing: 2006 100%, Least missing: 2018 97%
```

Warning: Removed 95909 rows containing non-finite values ('stat_count()').

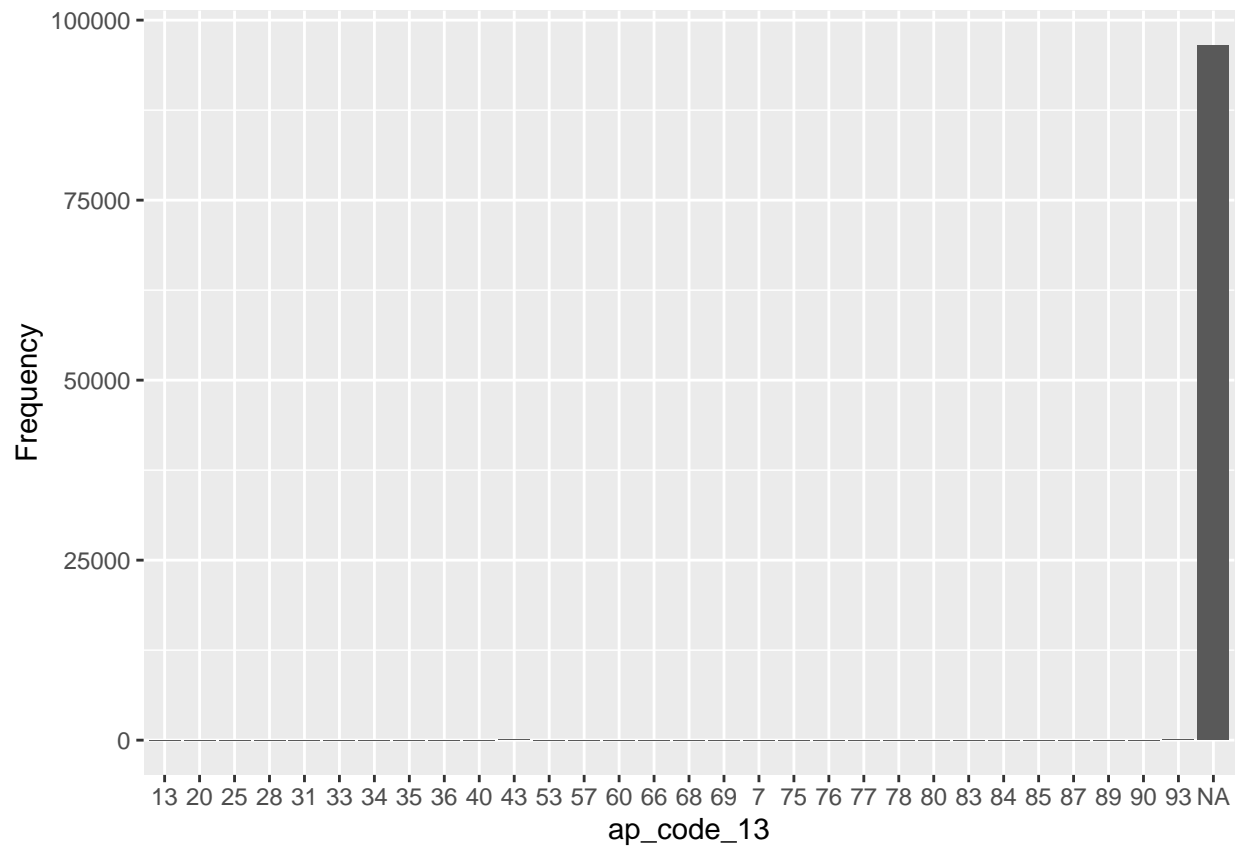


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_12, type: numeric, requested: yes
[1] Values (10 unique): NA, 0, 1, 5, 2, ...
[1] Missing: 99.7%
```

Warning: Removed 96737 rows containing non-finite values ('stat_count()').



```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_13, type: character, requested: yes
[1] Values (31 unique): NA, 93, 66, 90, 25, ...
[1] Missing: 99.4%
```



```
[1] is used in feature engineering
```

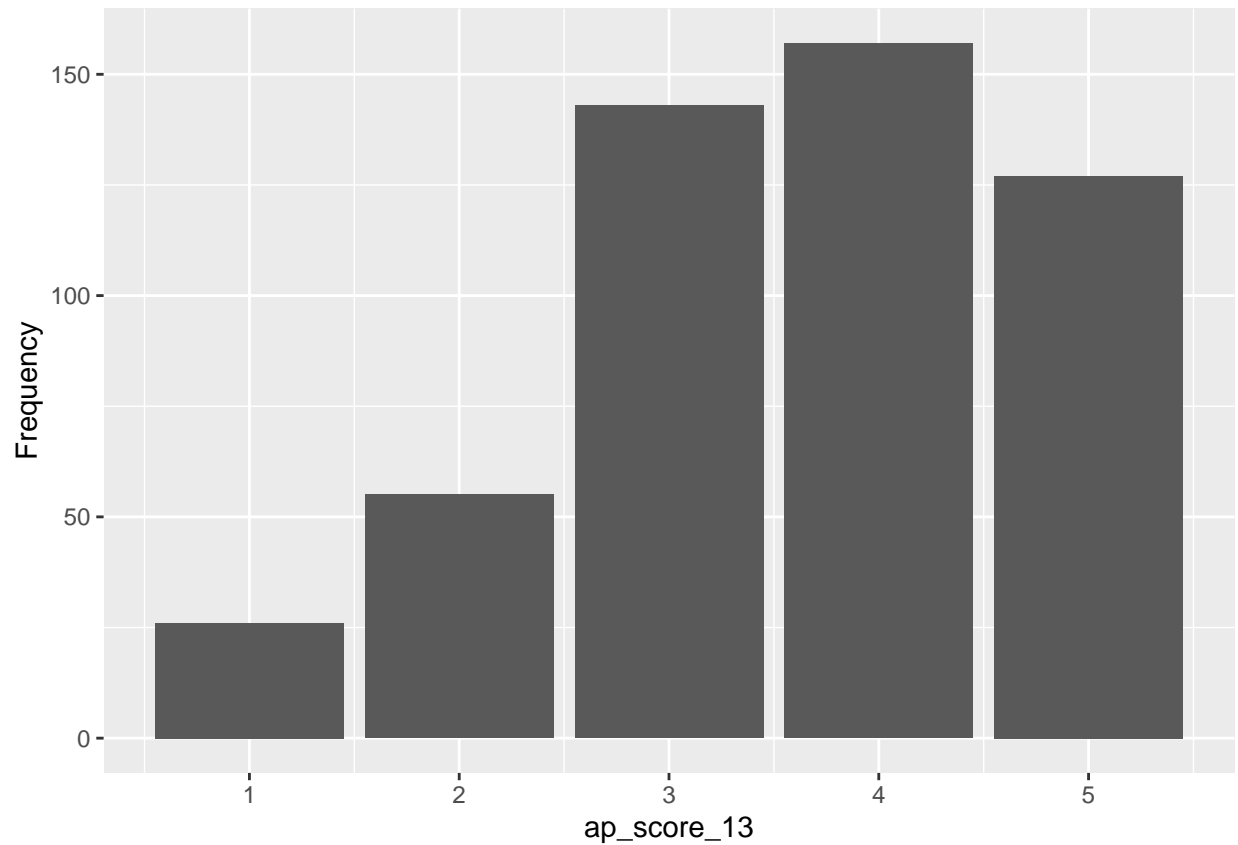
```
[1] -----
```

```
[1] Variable: ap_score_13, type: numeric, requested: yes
```

```
[1] Values (6 unique): NA, 3, 5, 4, 1, ...
```

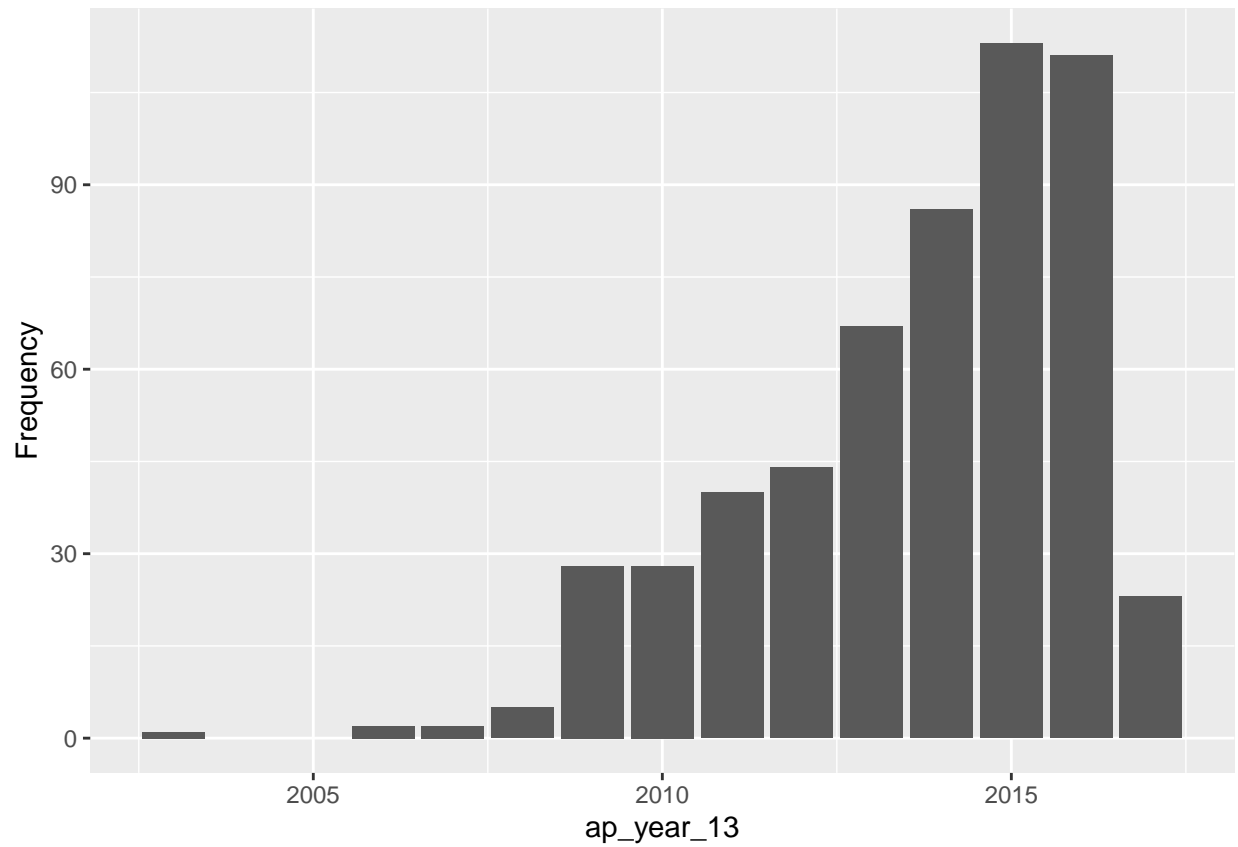
```
[1] Missing: 99.5%
```

```
Warning: Removed 96541 rows containing non-finite values ('stat_count()').
```



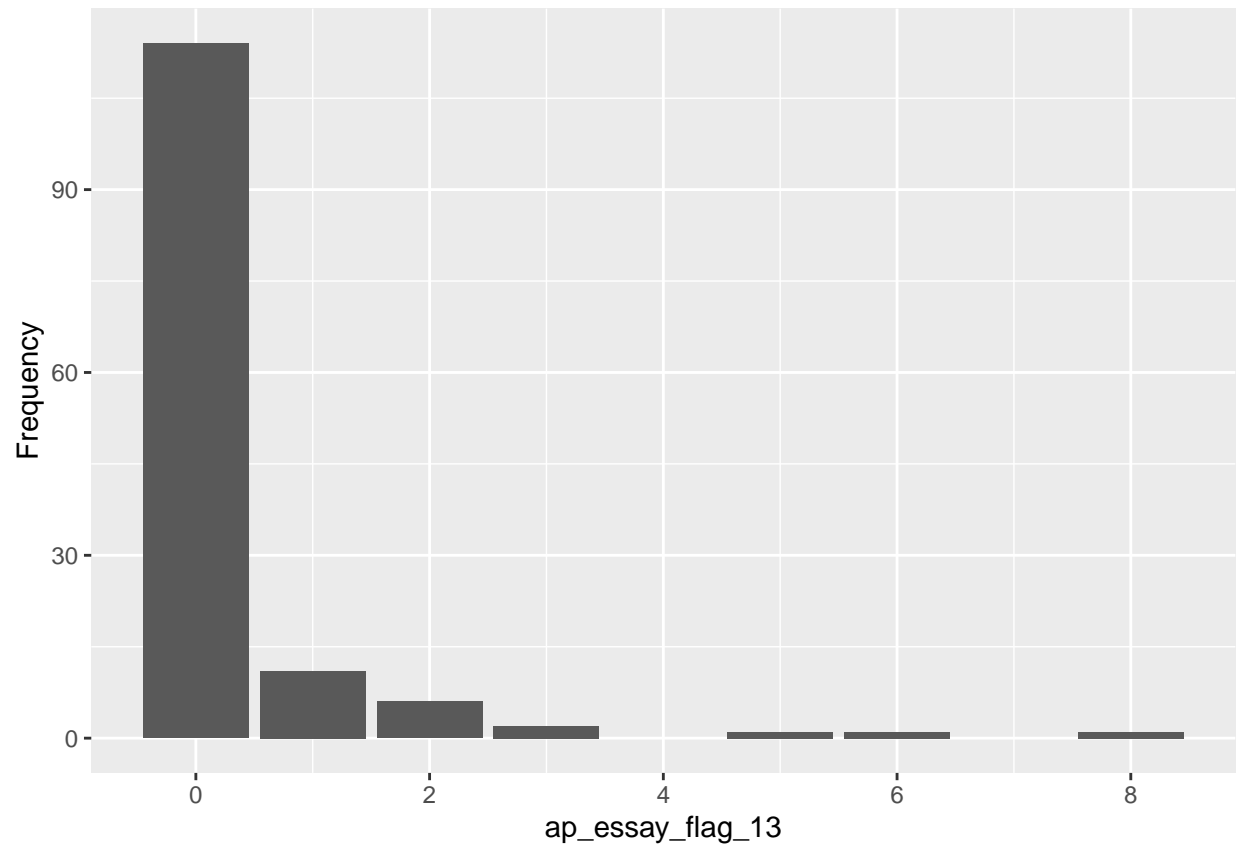
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_13, type: integer, requested: yes
[1] Values (14 unique): NA, 2015, 2017, 2016, 2013, ...
[1] Missing: 99.4%
```

Warning: Removed 96499 rows containing non-finite values ('stat_count()').

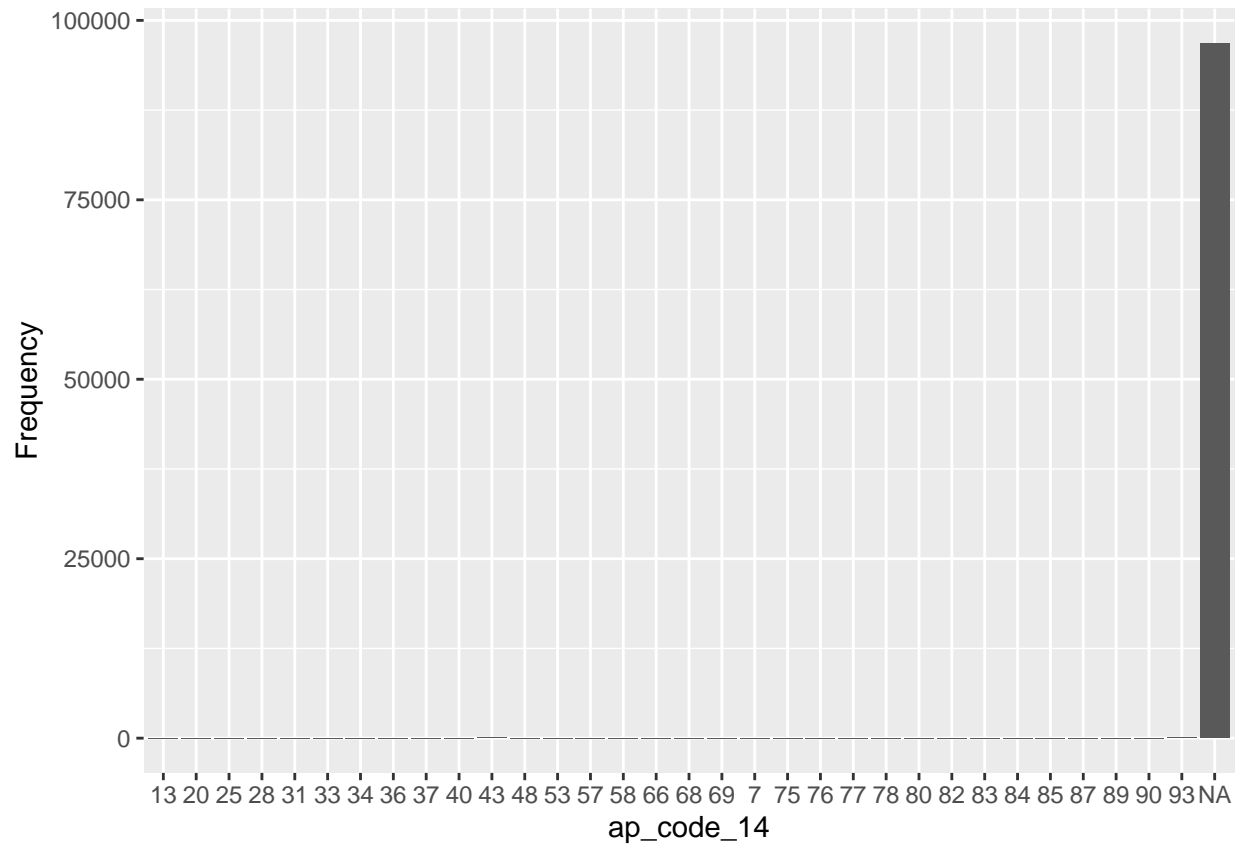


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_13, type: numeric, requested: yes
[1] Values (8 unique): NA, 0, 2, 1, 5, ...
[1] Missing: 99.9%
```

Warning: Removed 96913 rows containing non-finite values ('stat_count()').



```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_14, type: character, requested: yes
[1] Values (33 unique): NA, 83, 20, 93, 87, ...
[1] Missing: 99.7%
```



```
[1] is used in feature engineering
```

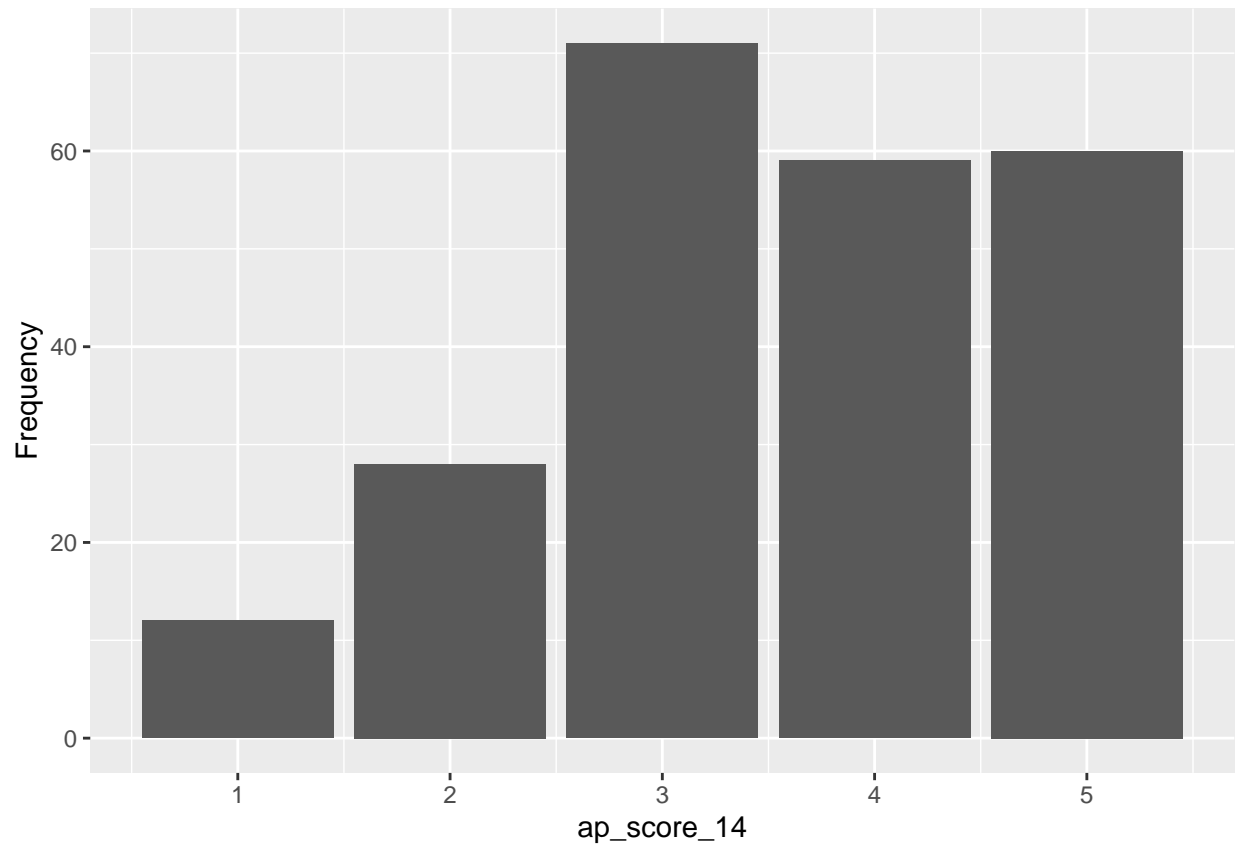
```
[1] -----
```

```
[1] Variable: ap_score_14, type: numeric, requested: yes
```

```
[1] Values (6 unique): NA, 5, 4, 3, 1, ...
```

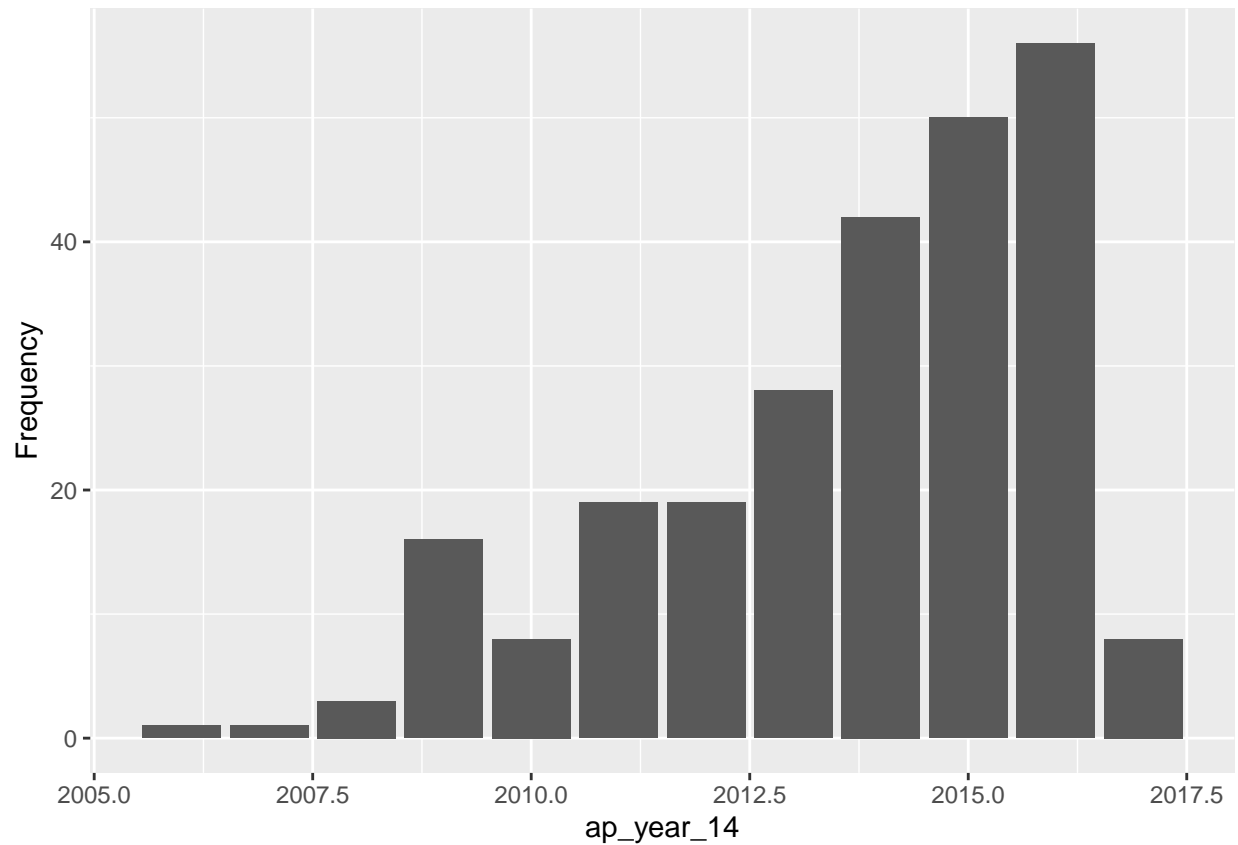
```
[1] Missing: 99.8%
```

```
Warning: Removed 96819 rows containing non-finite values ('stat_count()').
```



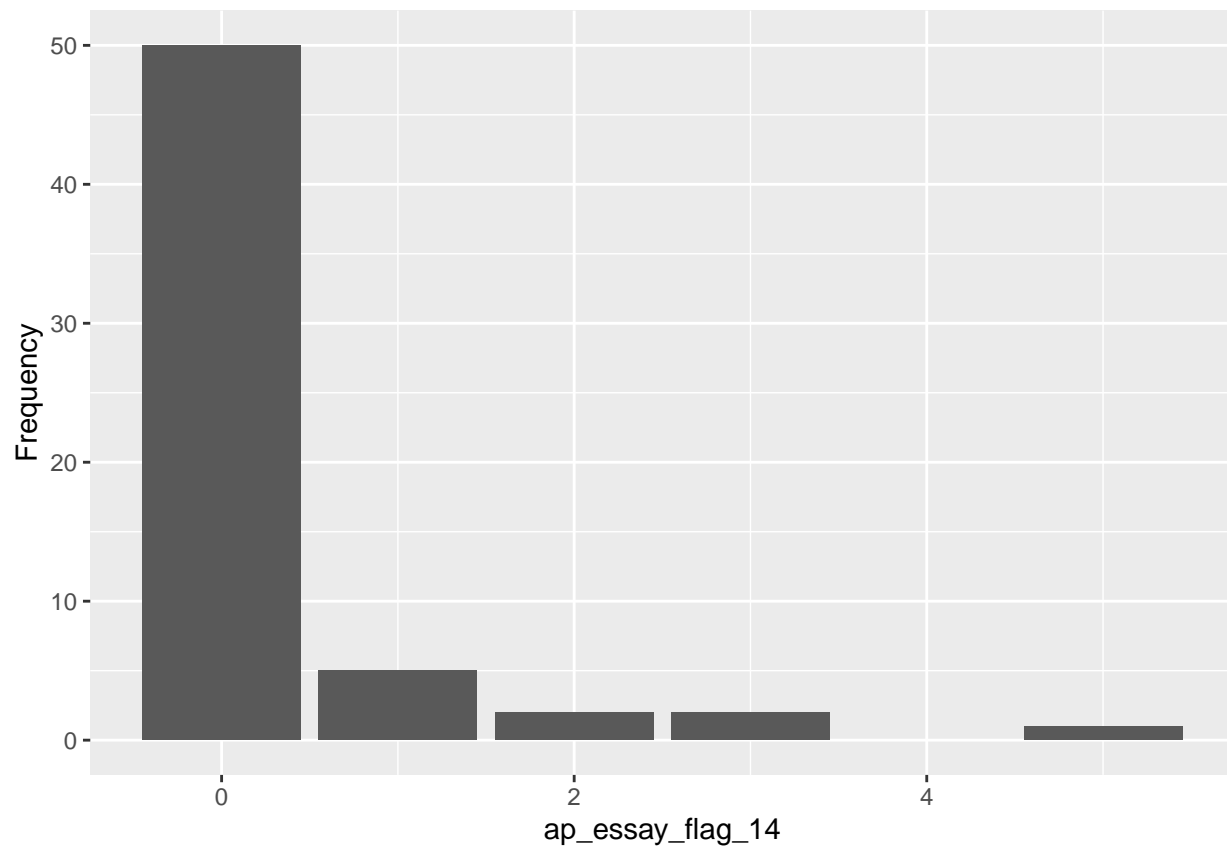
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_14, type: integer, requested: yes
[1] Values (13 unique): NA, 2017, 2016, 2015, 2013, ...
[1] Missing: 99.7%
```

Warning: Removed 96798 rows containing non-finite values ('stat_count()').

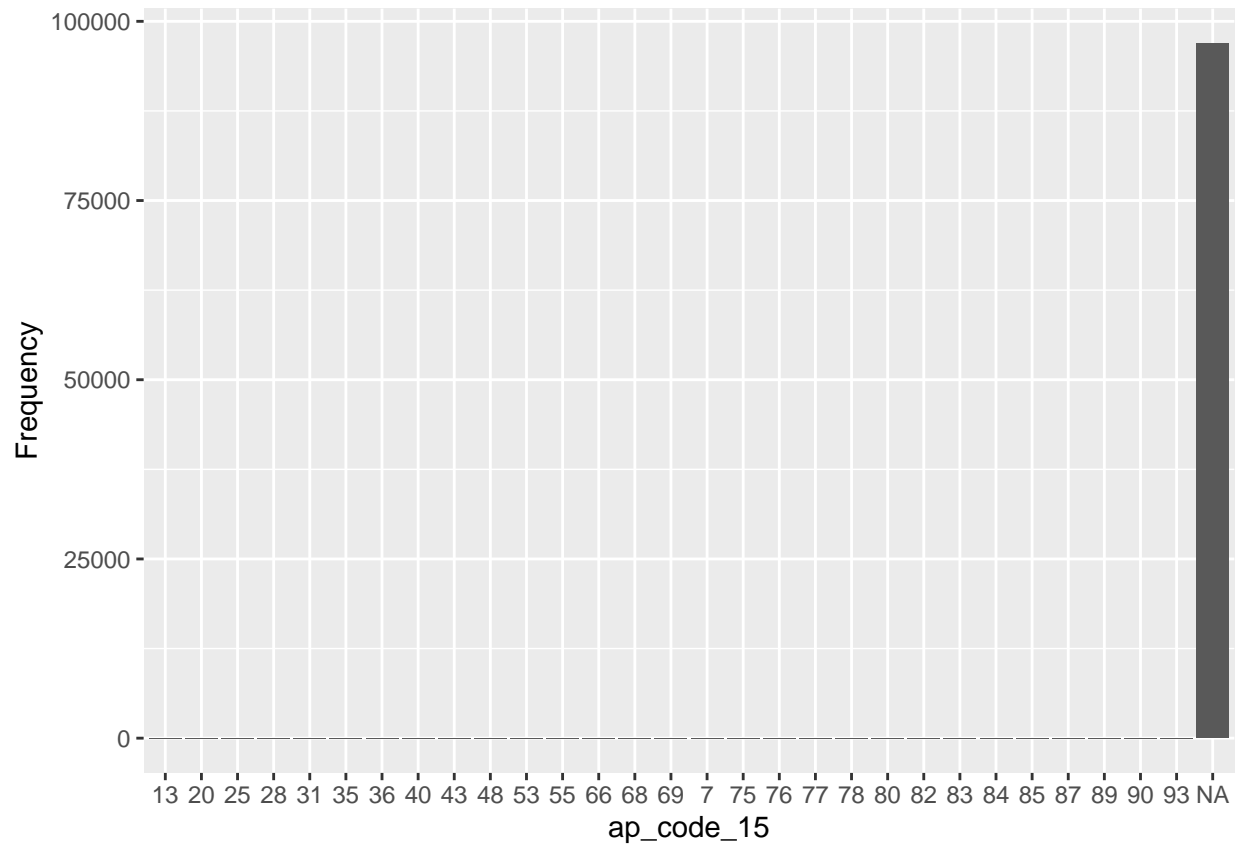


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_14, type: numeric, requested: yes
[1] Values (6 unique): NA, 0, 1, 2, 5, ...
[1] Missing: 99.9%
```

Warning: Removed 96989 rows containing non-finite values ('stat_count()').

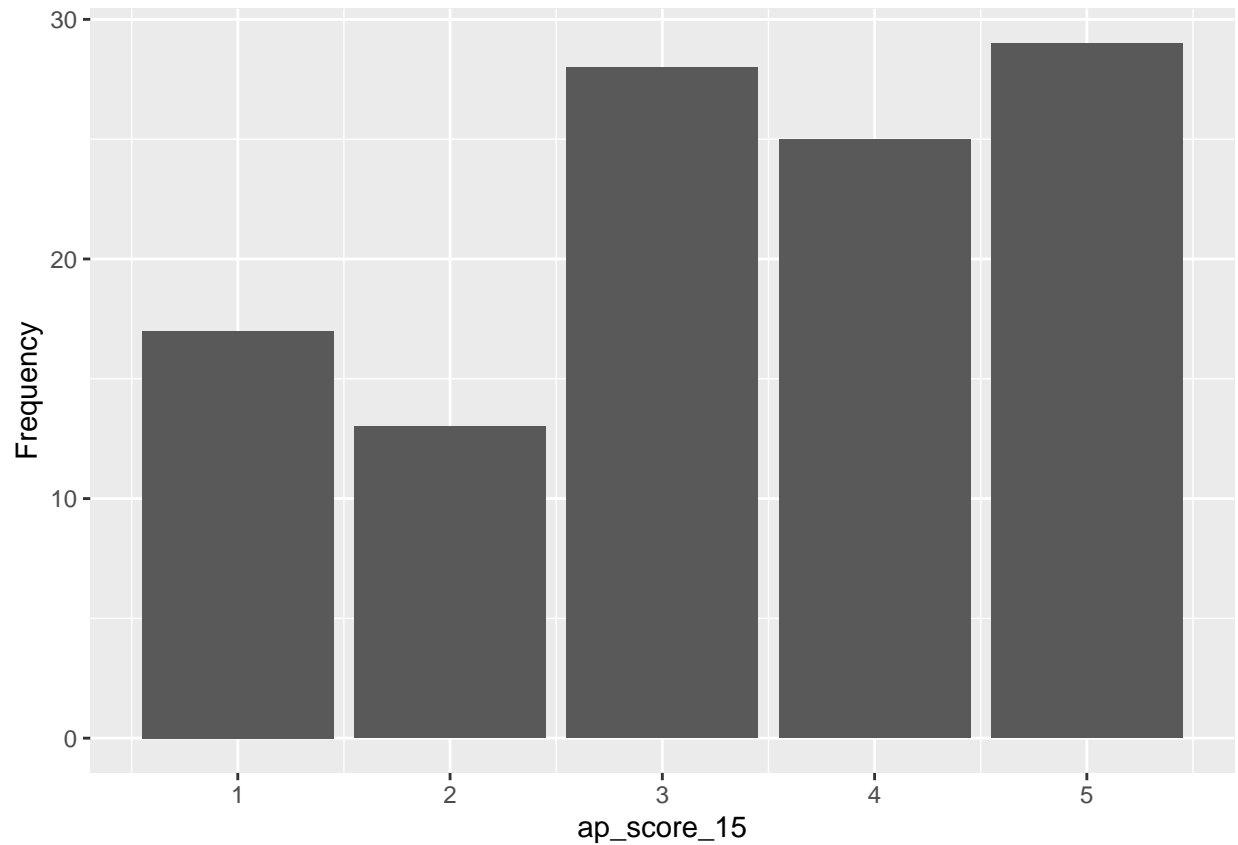


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_15, type: character, requested: yes
[1] Values (30 unique): NA, 93, 68, 20, 76, ...
[1] Missing: 99.9%
```



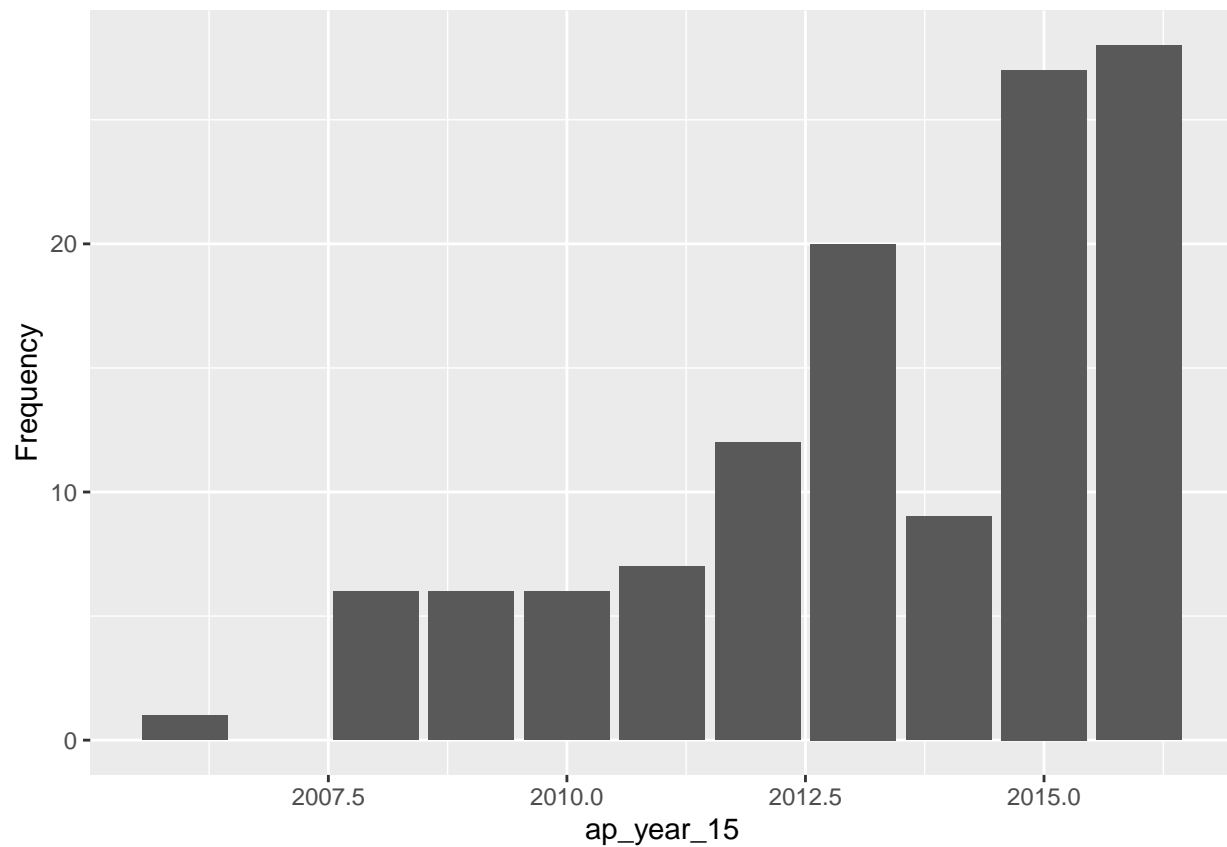
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_15, type: numeric, requested: yes
[1] Values (6 unique): NA, 3, 5, 1, 2, ...
[1] Missing: 99.9%
```

Warning: Removed 96937 rows containing non-finite values ('stat_count()').



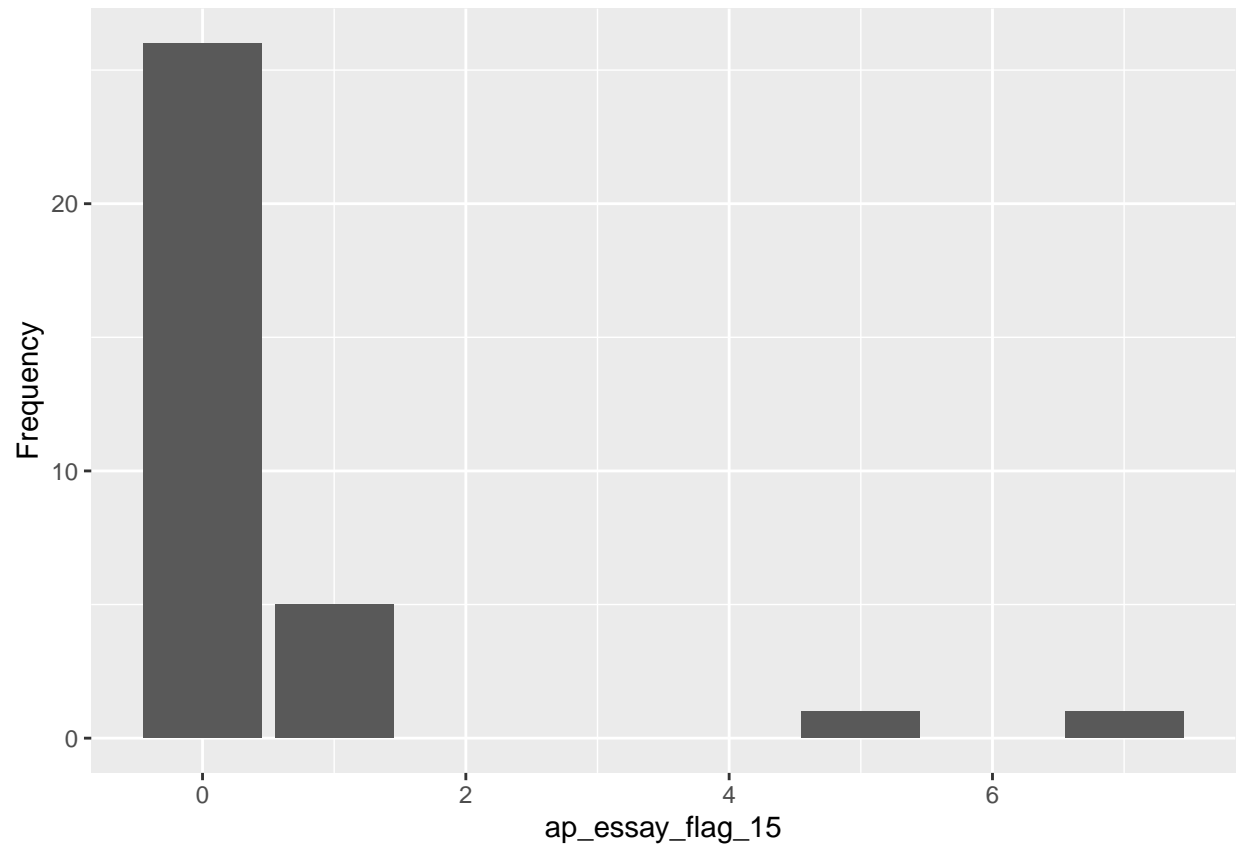
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_15, type: integer, requested: yes
[1] Values (11 unique): NA, 2016, 2014, 2015, 2013, ...
[1] Missing: 99.9%
```

Warning: Removed 96927 rows containing non-finite values ('stat_count()').

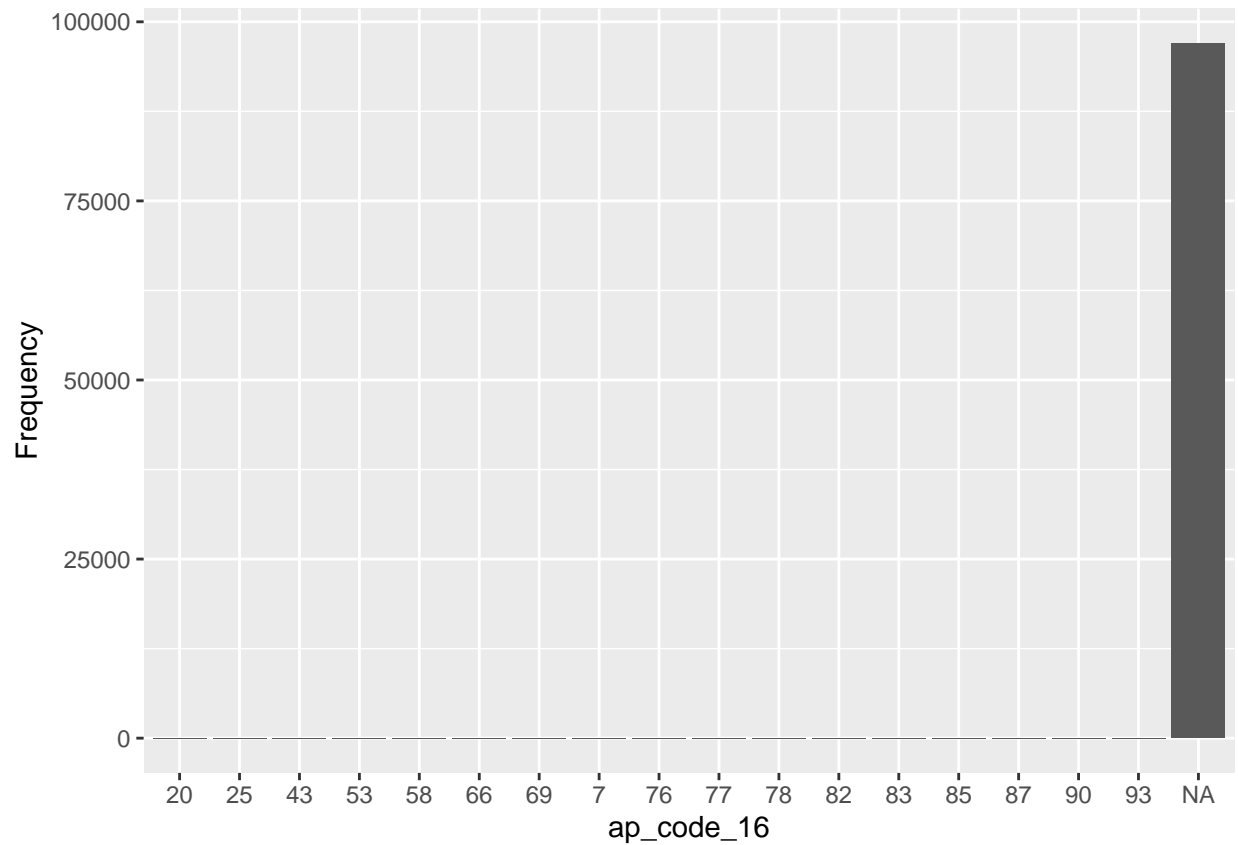


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_15, type: numeric, requested: yes
[1] Values (5 unique): NA, 0, 7, 5, 1
[1] Missing: 100%
```

Warning: Removed 97016 rows containing non-finite values ('stat_count()').

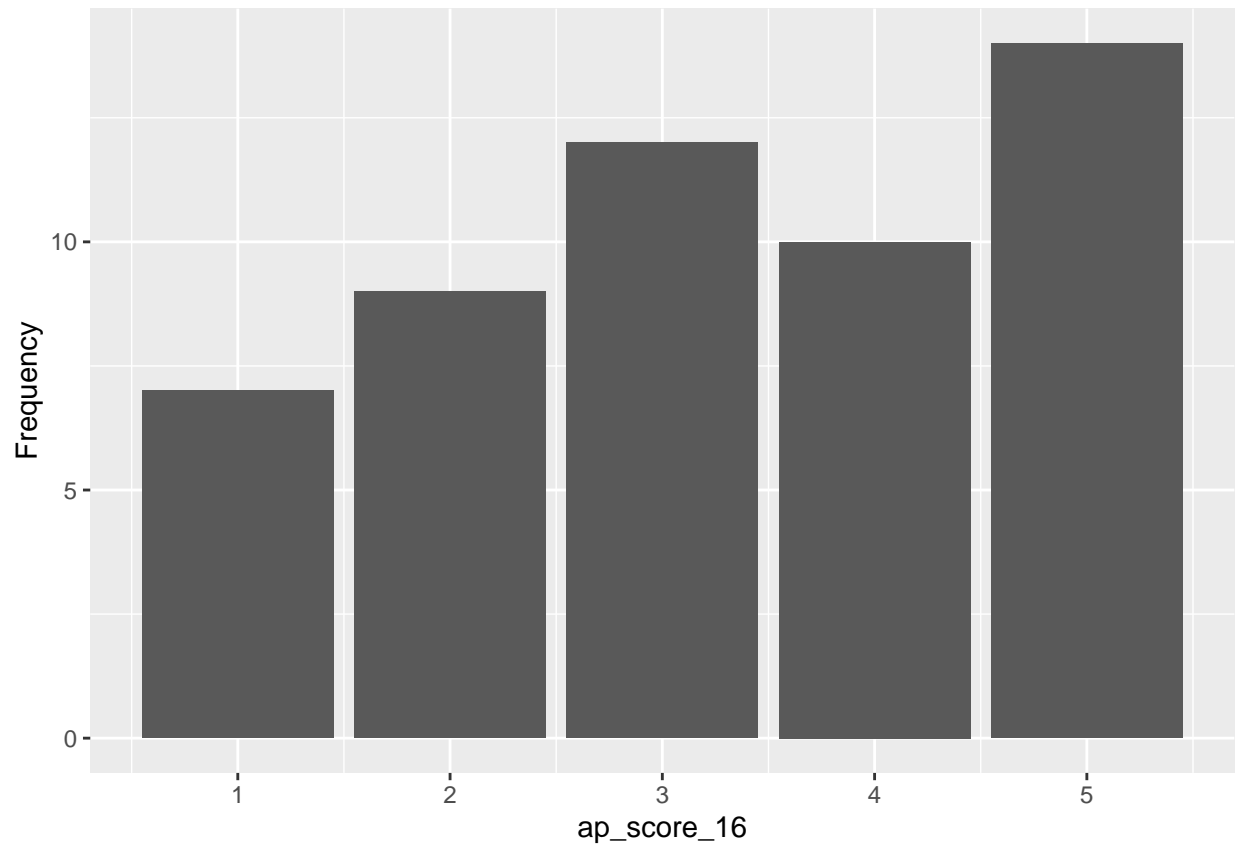


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_16, type: character, requested: yes
[1] Values (18 unique): NA, 69, 43, 77, 93, ...
[1] Missing: 99.9%
```



```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_16, type: numeric, requested: yes
[1] Values (6 unique): NA, 5, 4, 3, 2, ...
[1] Missing: 99.9%
```

Warning: Removed 96997 rows containing non-finite values ('stat_count()').



```
[1] is used in feature engineering
```

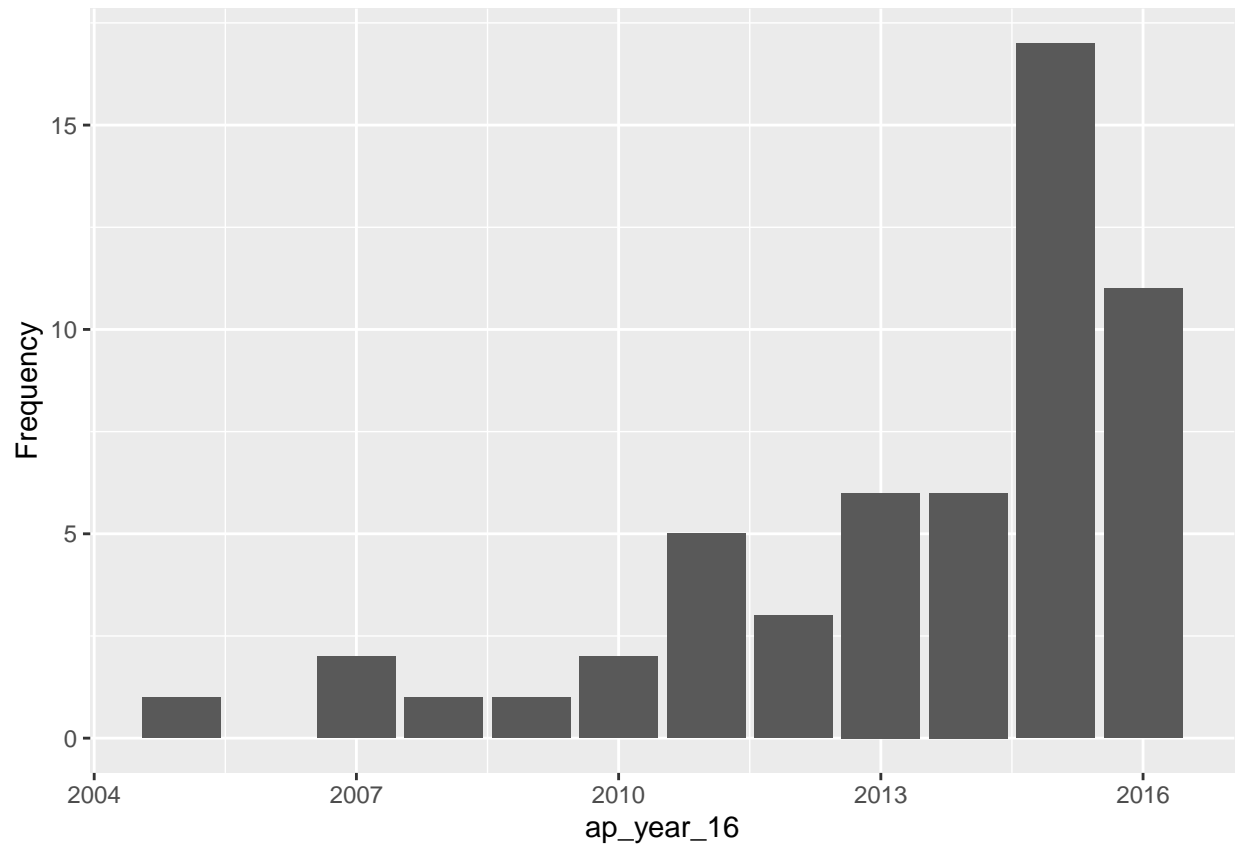
```
[1] -----
```

```
[1] Variable: ap_year_16, type: integer, requested: yes
```

```
[1] Values (12 unique): NA, 2016, 2015, 2013, 2014, ...
```

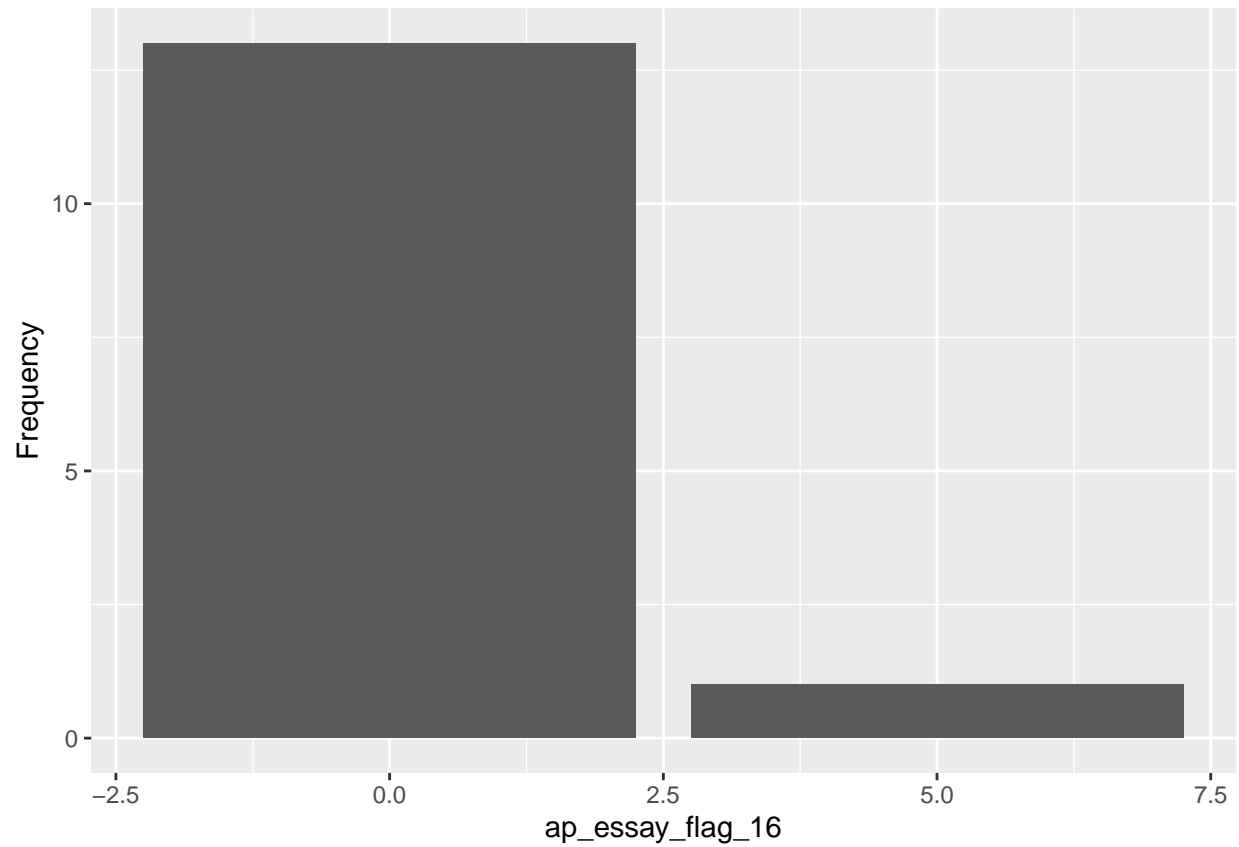
```
[1] Missing: 99.9%
```

```
Warning: Removed 96994 rows containing non-finite values ('stat_count()').
```

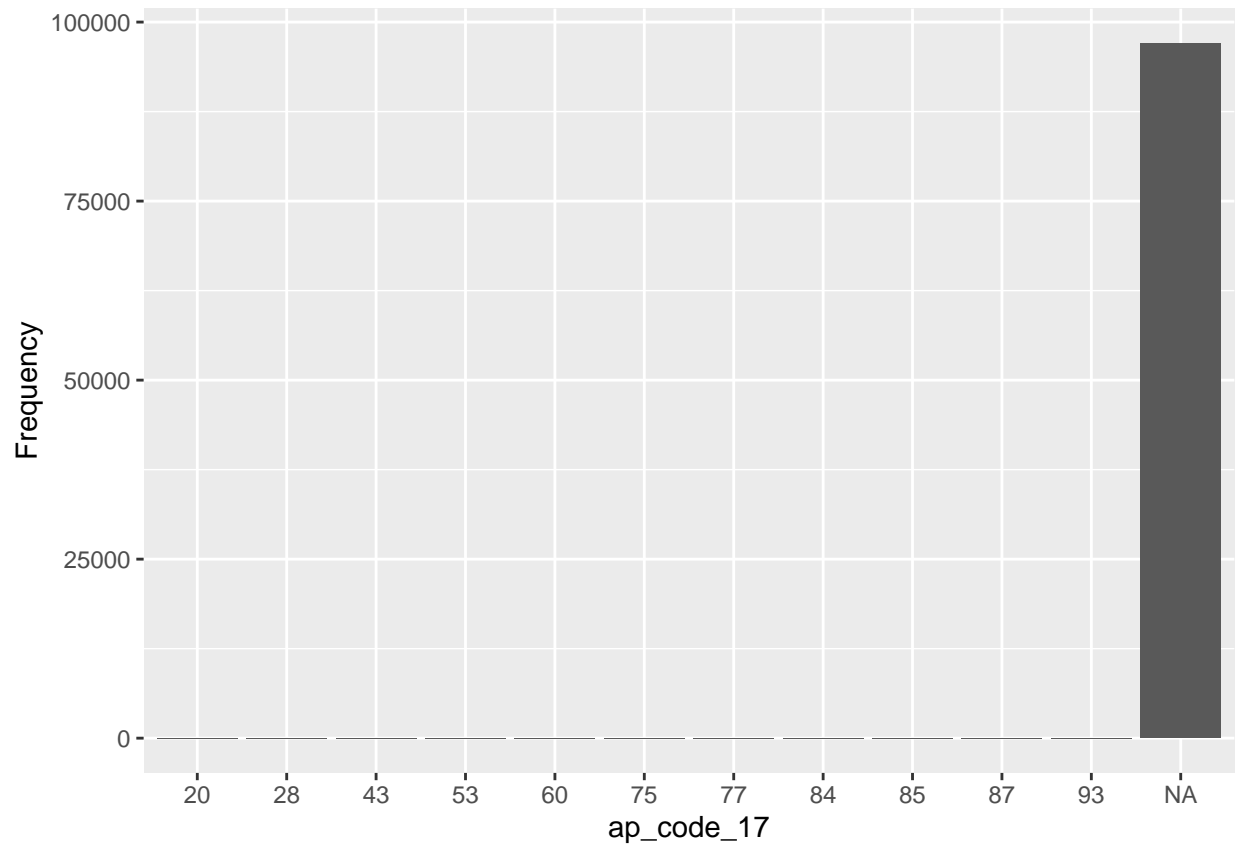


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_16, type: numeric, requested: yes
[1] Values (3 unique): NA, 0, 5
[1] Missing: 100%
```

Warning: Removed 97035 rows containing non-finite values ('stat_count()').

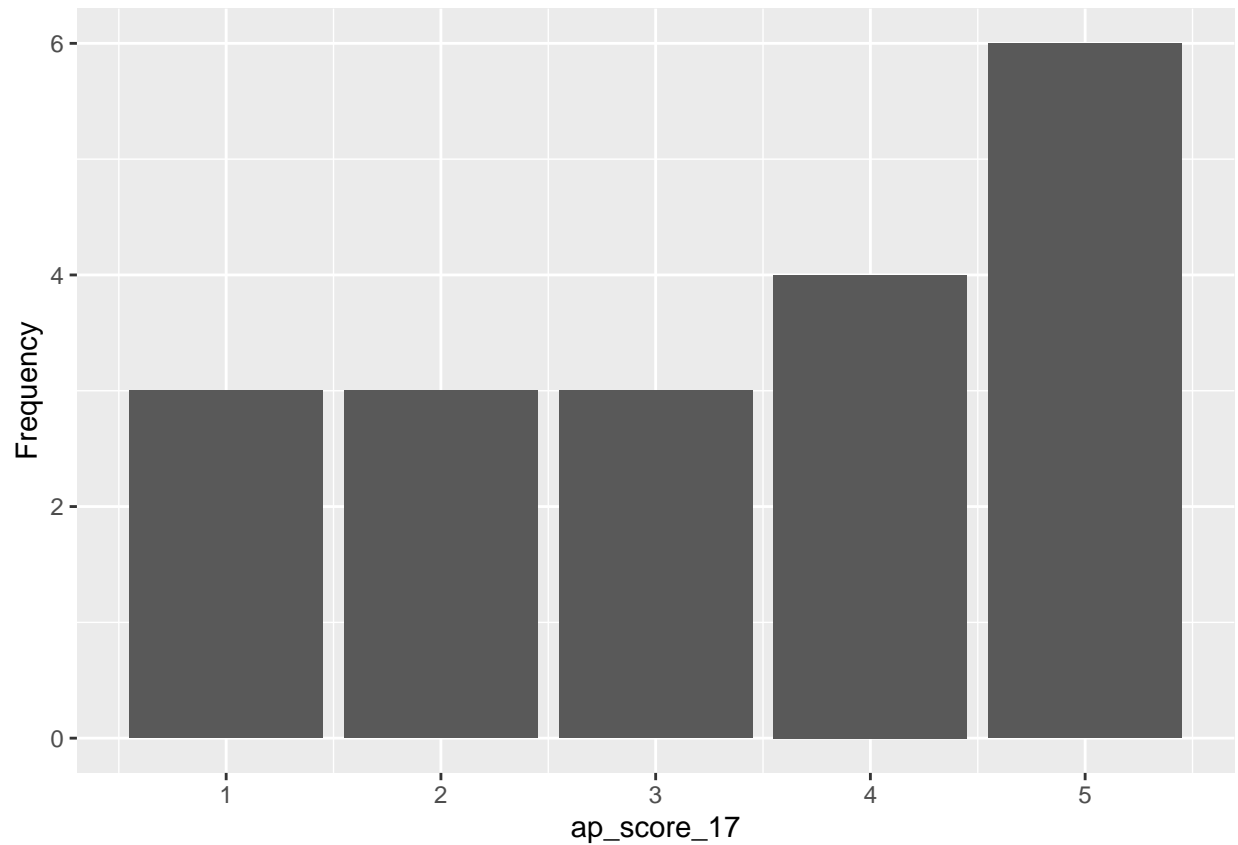


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_17, type: character, requested: yes
[1] Values (12 unique): NA, 93, 28, 87, 20, ...
[1] Missing: 100%
```



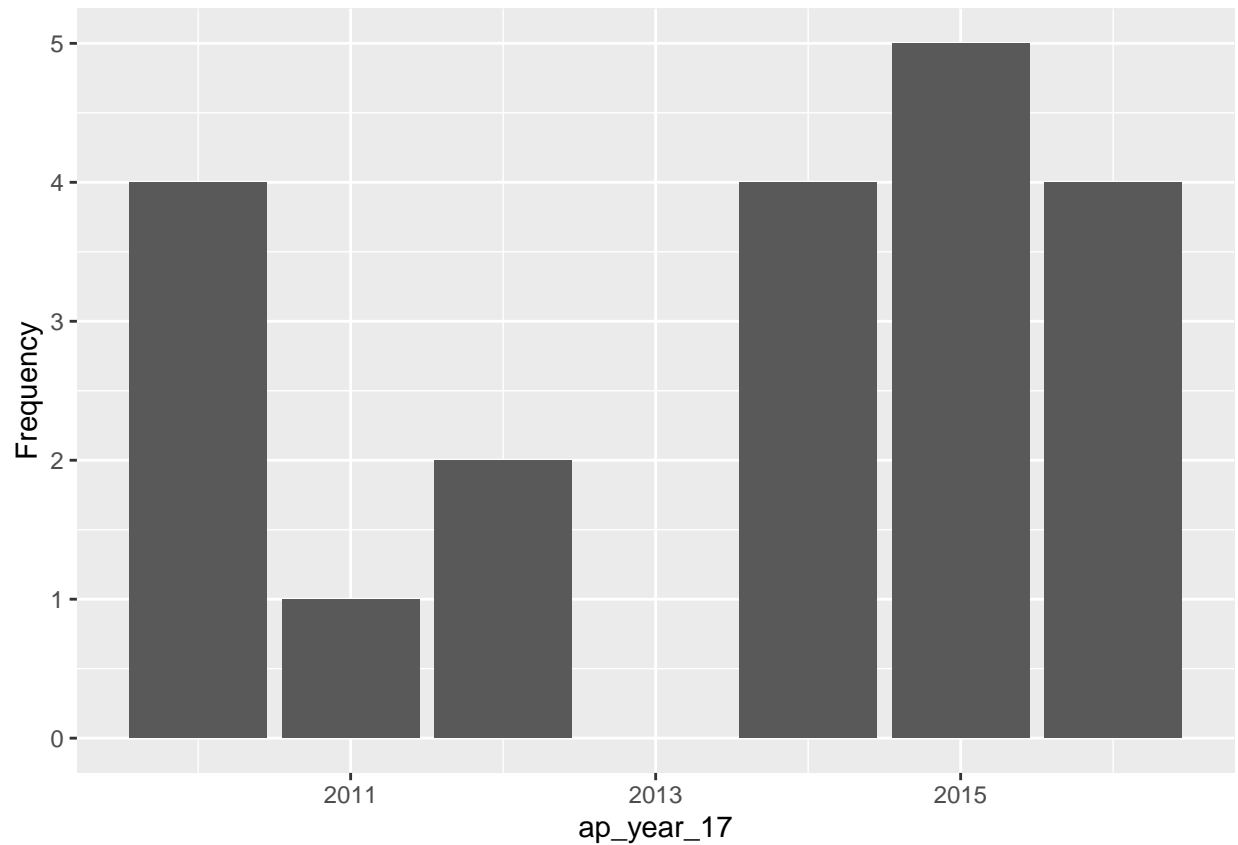
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_17, type: numeric, requested: yes
[1] Values (6 unique): NA, 3, 5, 4, 1, ...
[1] Missing: 100%
```

Warning: Removed 97030 rows containing non-finite values ('stat_count()').



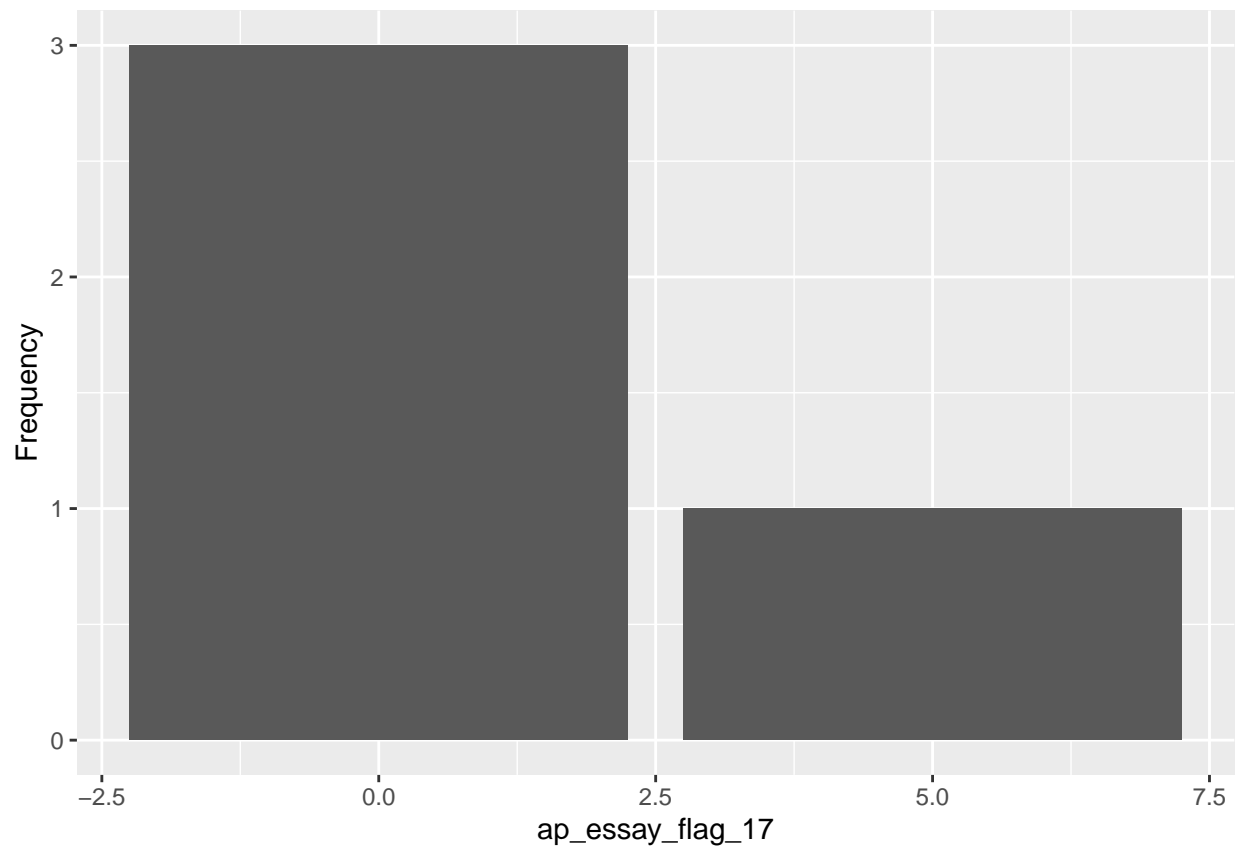
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_17, type: integer, requested: yes
[1] Values (7 unique): NA, 2016, 2015, 2012, 2014, ...
[1] Missing: 100%
```

Warning: Removed 97029 rows containing non-finite values ('stat_count()').

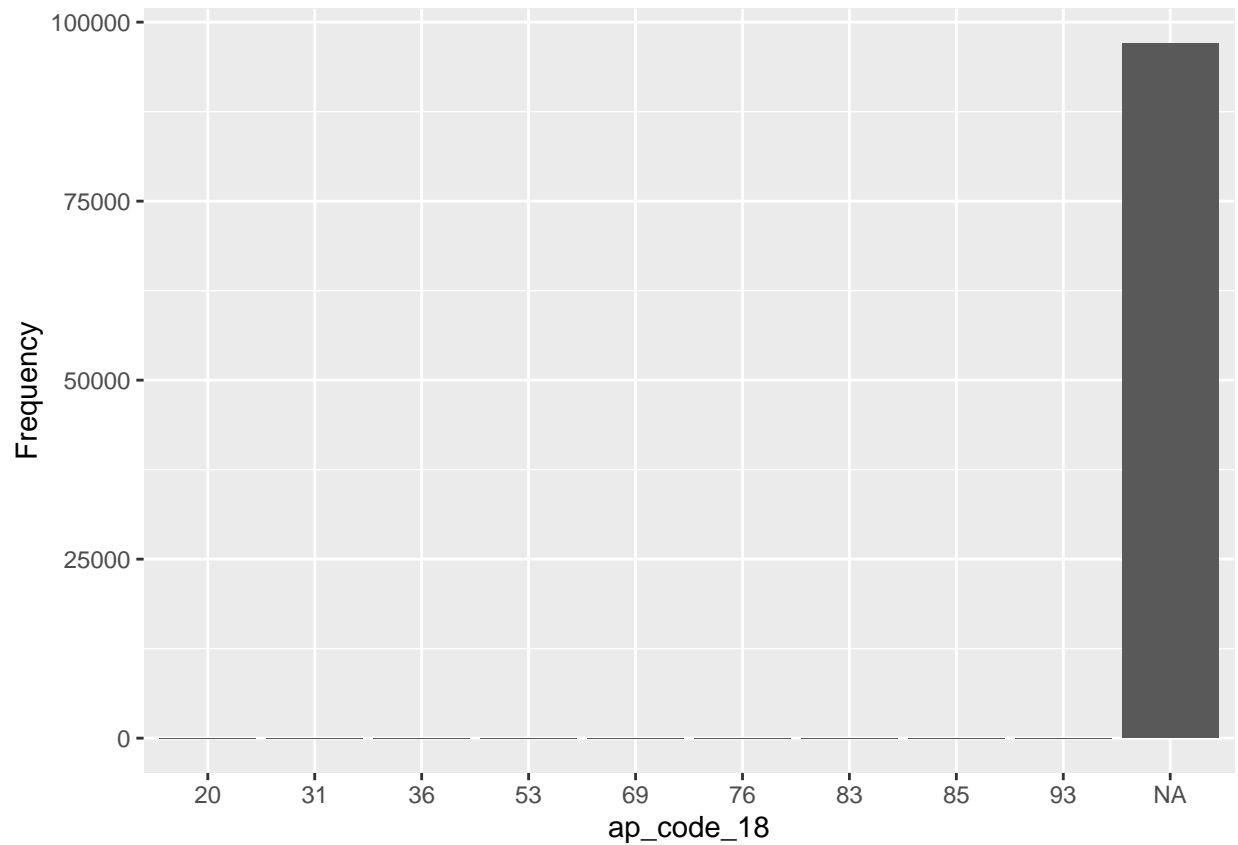


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_17, type: numeric, requested: yes
[1] Values (3 unique): NA, 5, 0
[1] Missing: 100%
```

Warning: Removed 97045 rows containing non-finite values ('stat_count()').

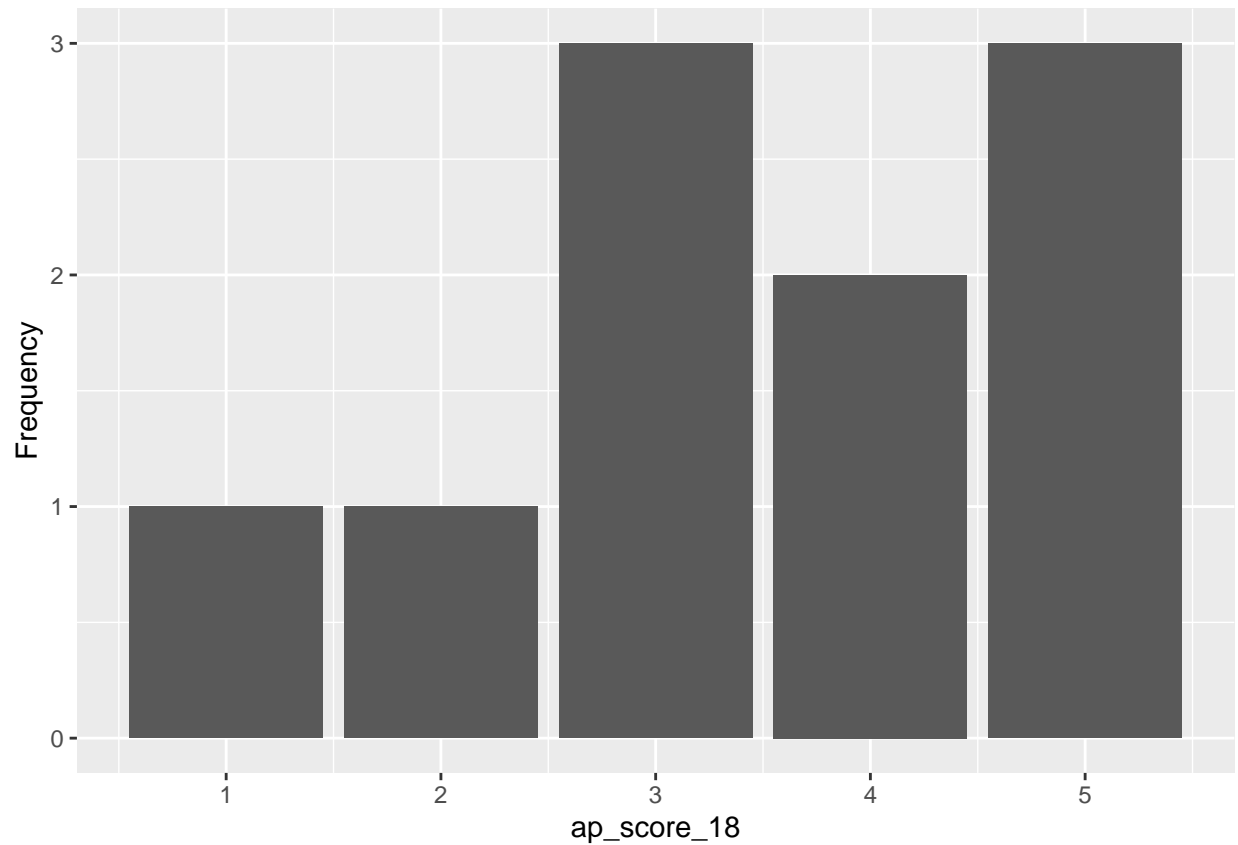


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_18, type: character, requested: yes
[1] Values (10 unique): NA, 20, 93, 53, 85, ...
[1] Missing: 100%
```



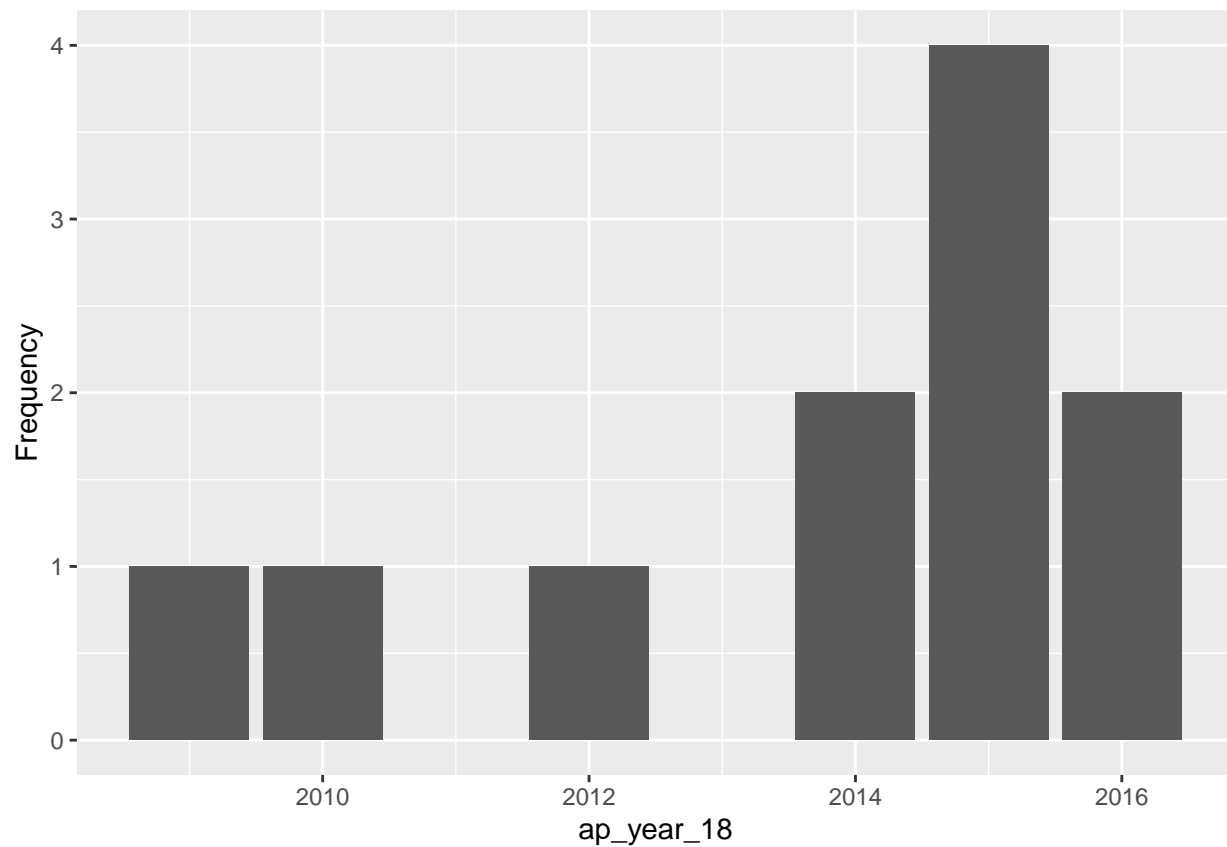
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_18, type: numeric, requested: yes
[1] Values (6 unique): NA, 4, 3, 5, 2, ...
[1] Missing: 100%
```

Warning: Removed 97039 rows containing non-finite values ('stat_count()').



```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_18, type: integer, requested: yes
[1] Values (7 unique): NA, 2012, 2014, 2015, 2016, ...
[1] Missing: 100%
```

Warning: Removed 97038 rows containing non-finite values ('stat_count()').



```
[1] is used in feature engineering
```

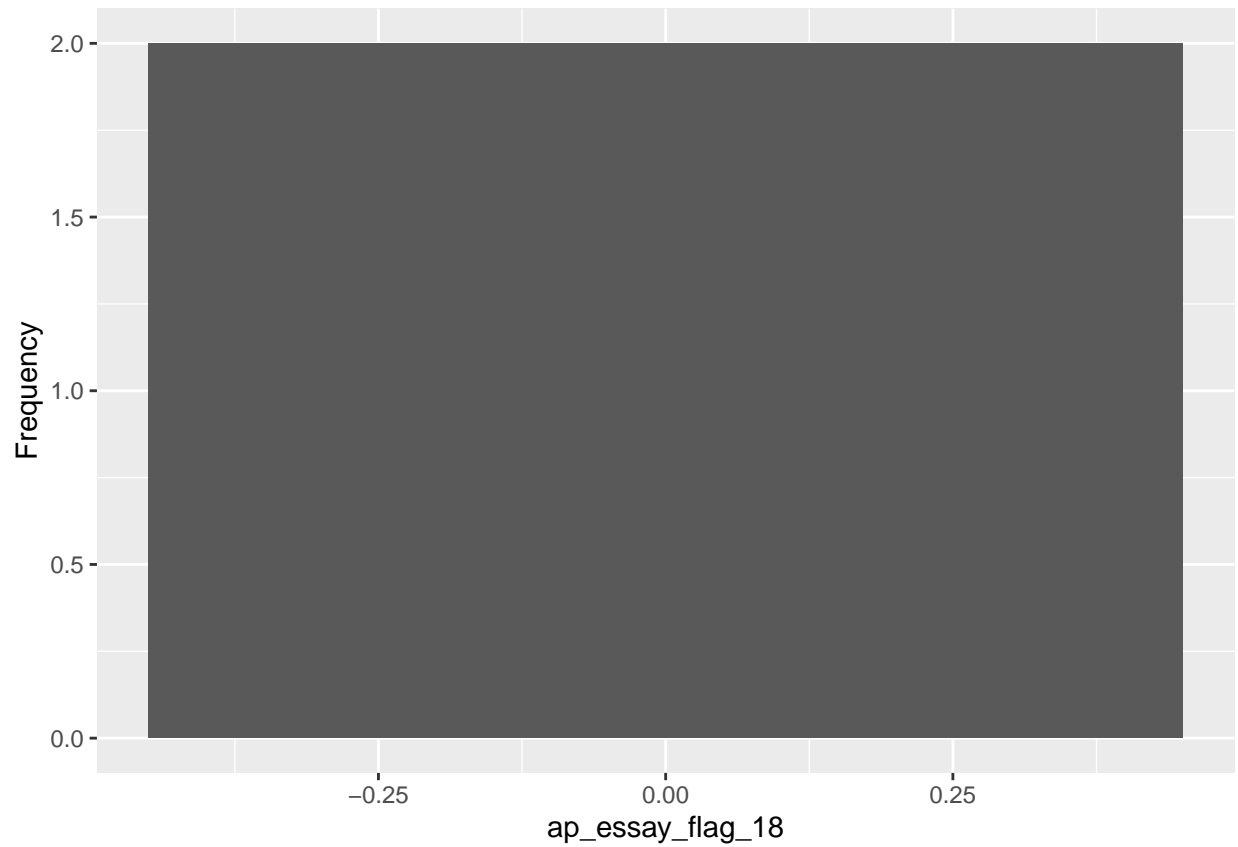
```
[1] -----
```

```
[1] Variable: ap_essay_flag_18, type: numeric, requested: yes
```

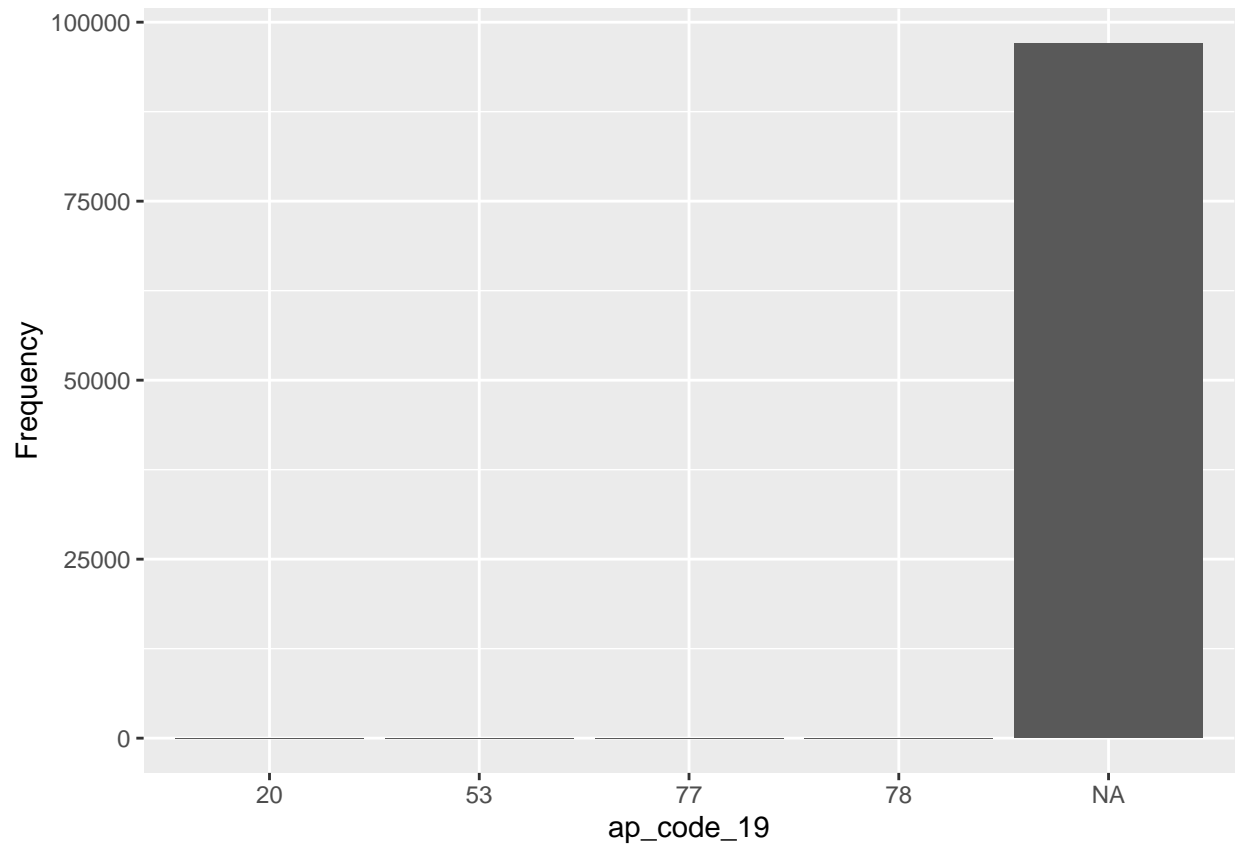
```
[1] Values (2 unique): NA, 0
```

```
[1] Missing: 100%
```

```
Warning: Removed 97047 rows containing non-finite values ('stat_count()').
```

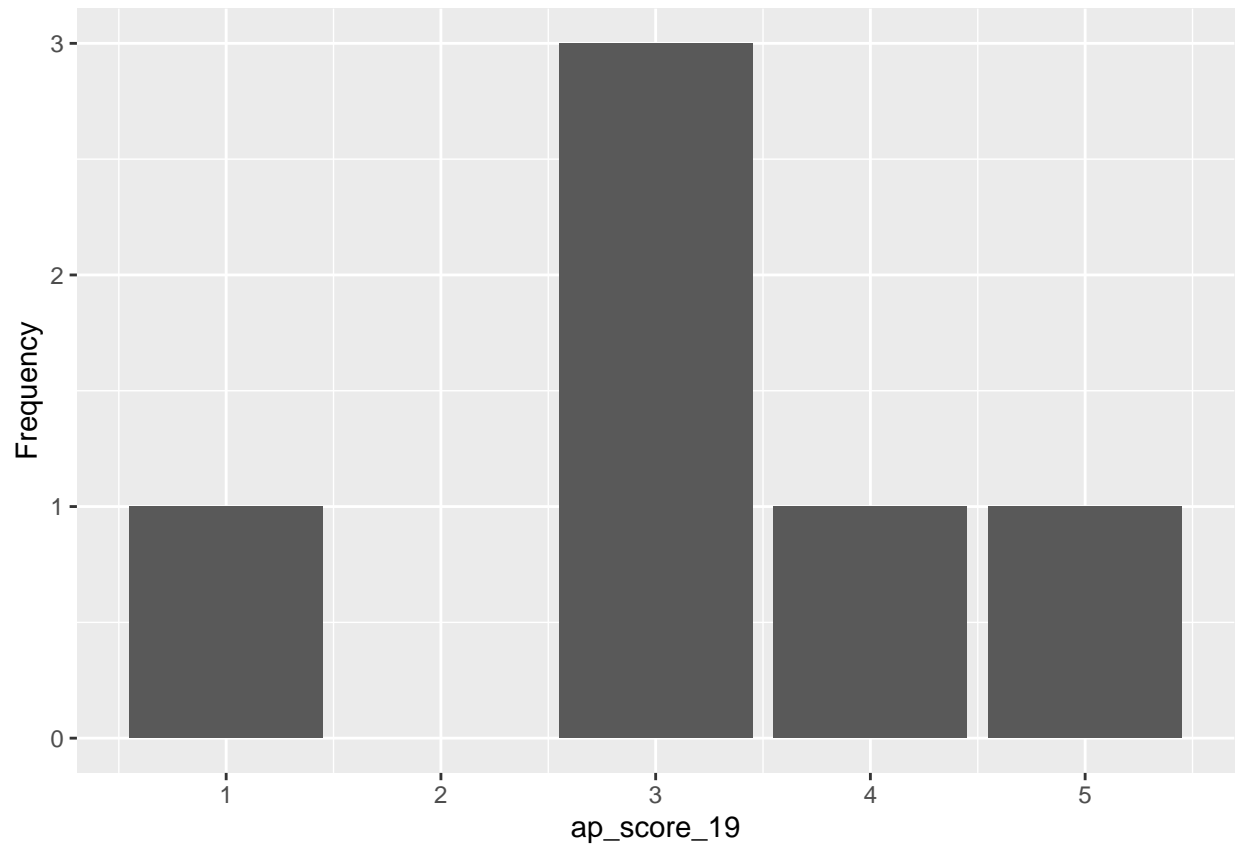



```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_19, type: character, requested: yes
[1] Values (5 unique): NA, 53, 20, 77, 78
[1] Missing: 100%
```



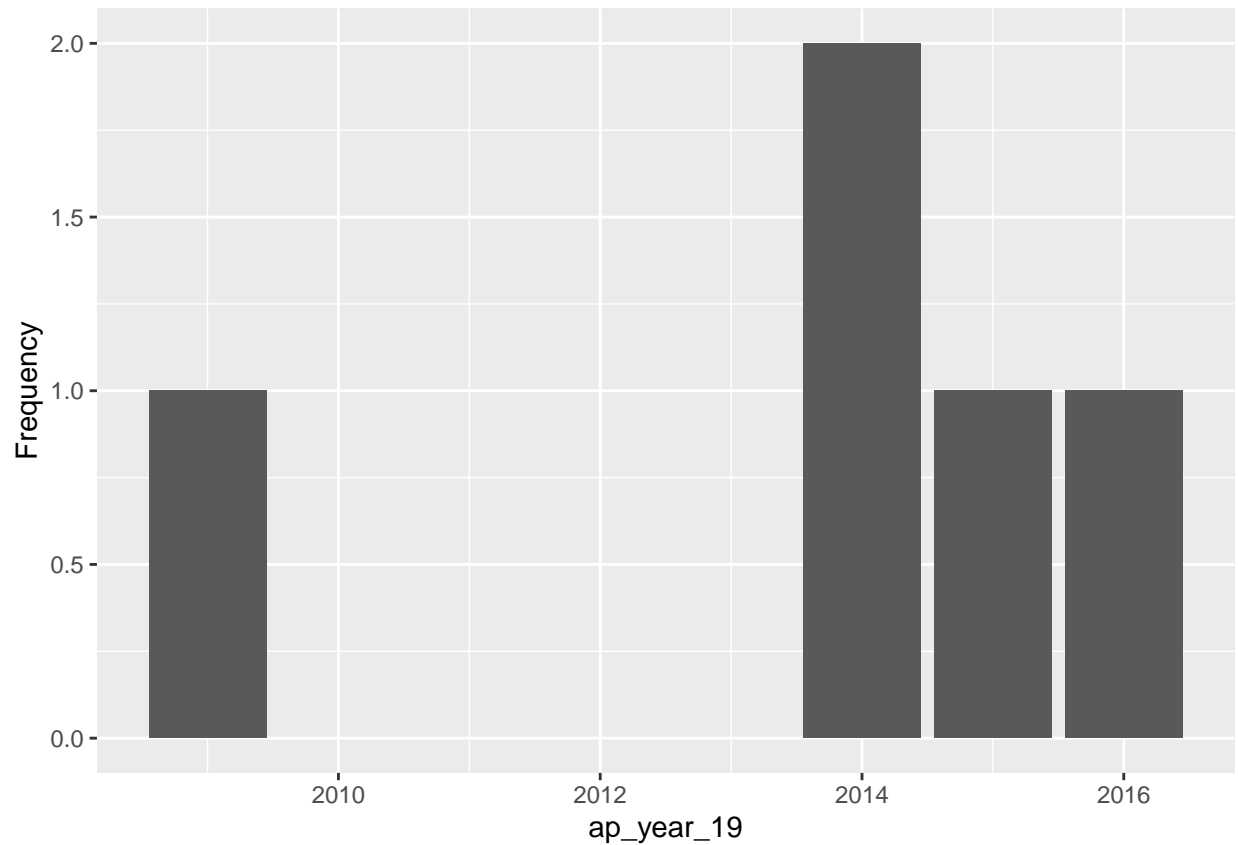
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_19, type: numeric, requested: yes
[1] Values (5 unique): NA, 3, 5, 4, 1
[1] Missing: 100%
```

Warning: Removed 97043 rows containing non-finite values ('stat_count()').



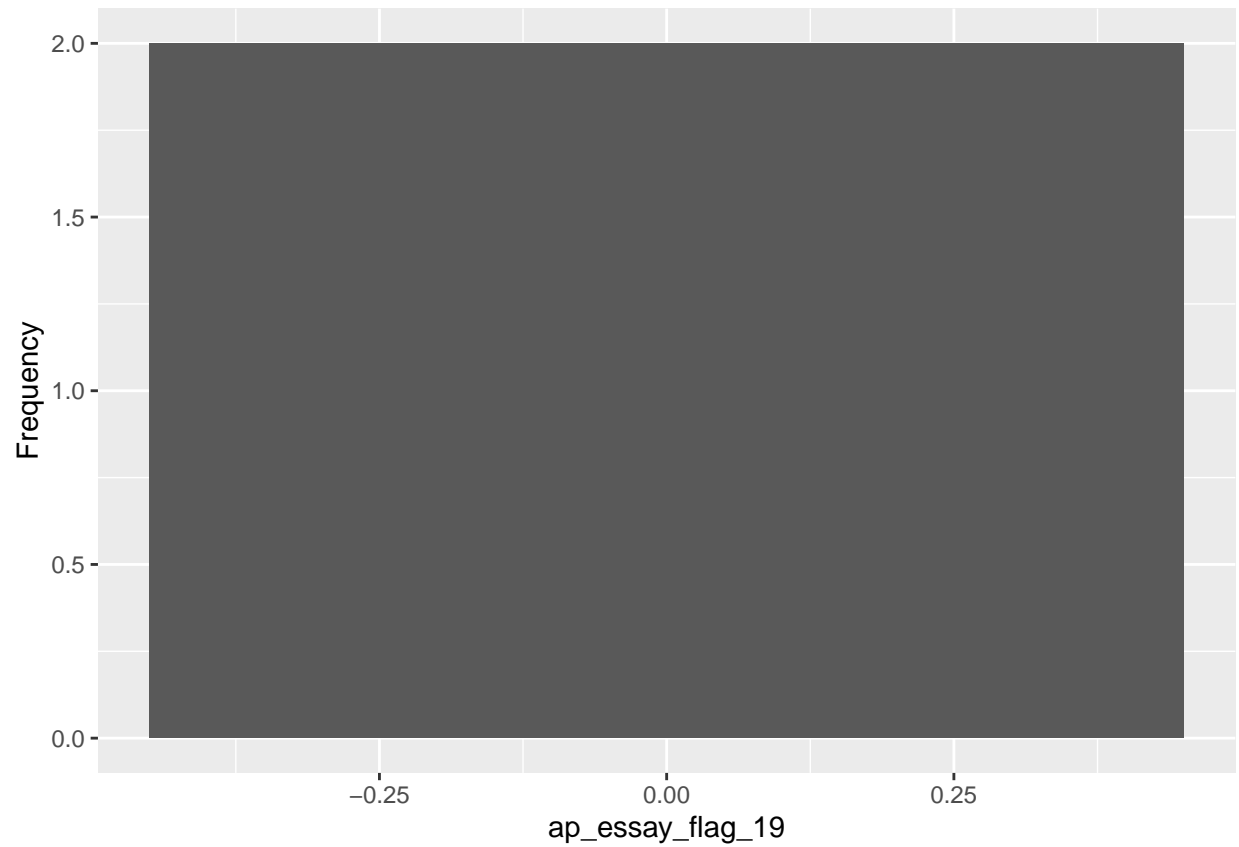
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_19, type: integer, requested: yes
[1] Values (5 unique): NA, 2014, 2015, 2016, 2009
[1] Missing: 100%
```

Warning: Removed 97044 rows containing non-finite values ('stat_count()').

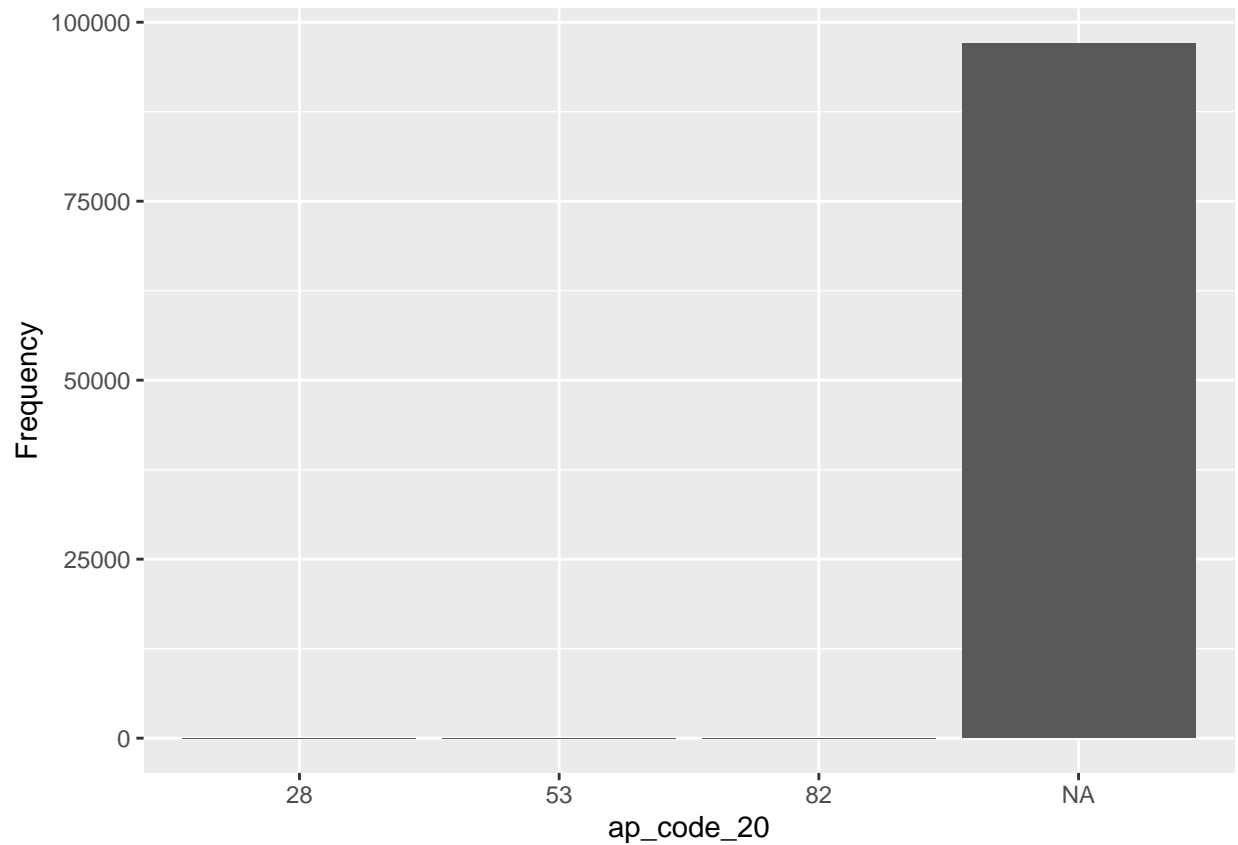


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_19, type: numeric, requested: yes
[1] Values (2 unique): NA, 0
[1] Missing: 100%
```

Warning: Removed 97047 rows containing non-finite values ('stat_count()').

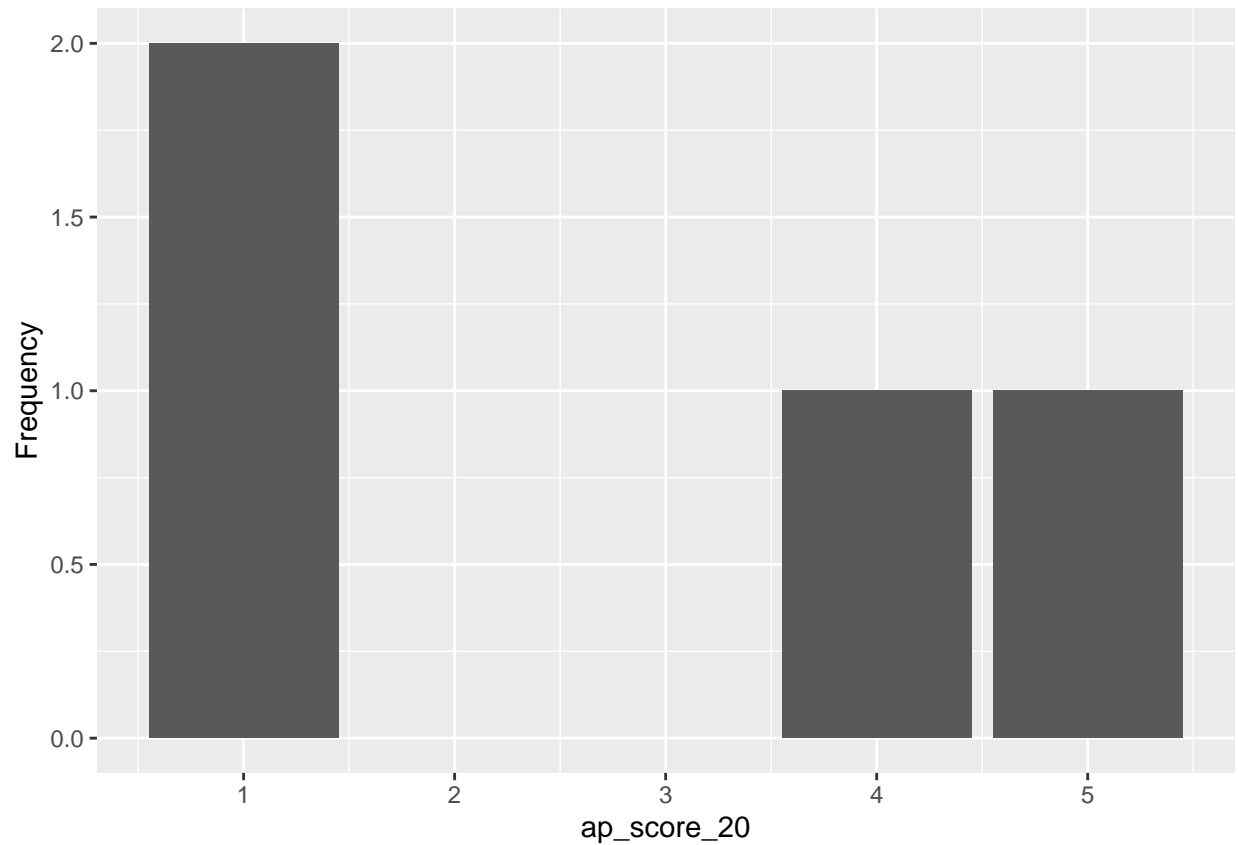


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code_20, type: character, requested: yes
[1] Values (4 unique): NA, 53, 28, 82
[1] Missing: 100%
```



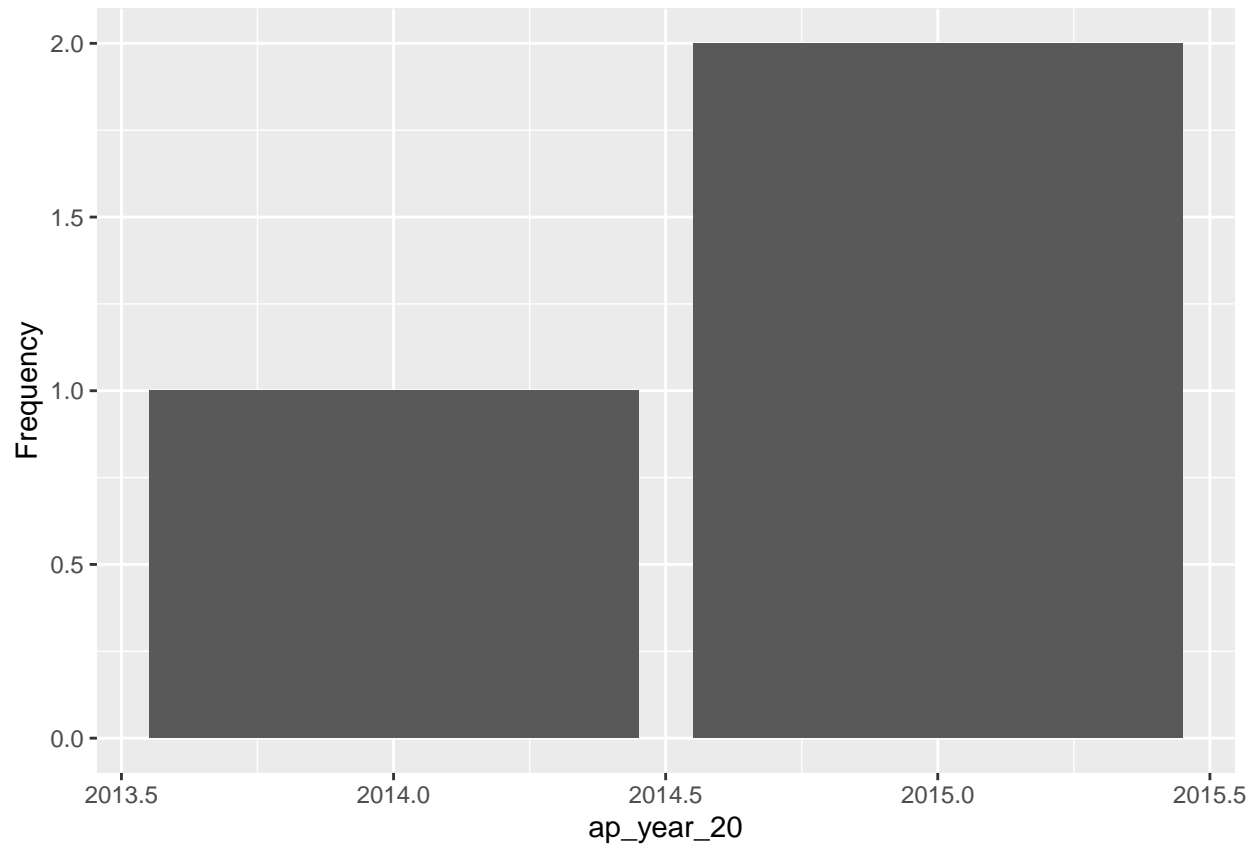
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_score_20, type: numeric, requested: yes
[1] Values (4 unique): NA, 5, 4, 1
[1] Missing: 100%
```

Warning: Removed 97045 rows containing non-finite values ('stat_count()').



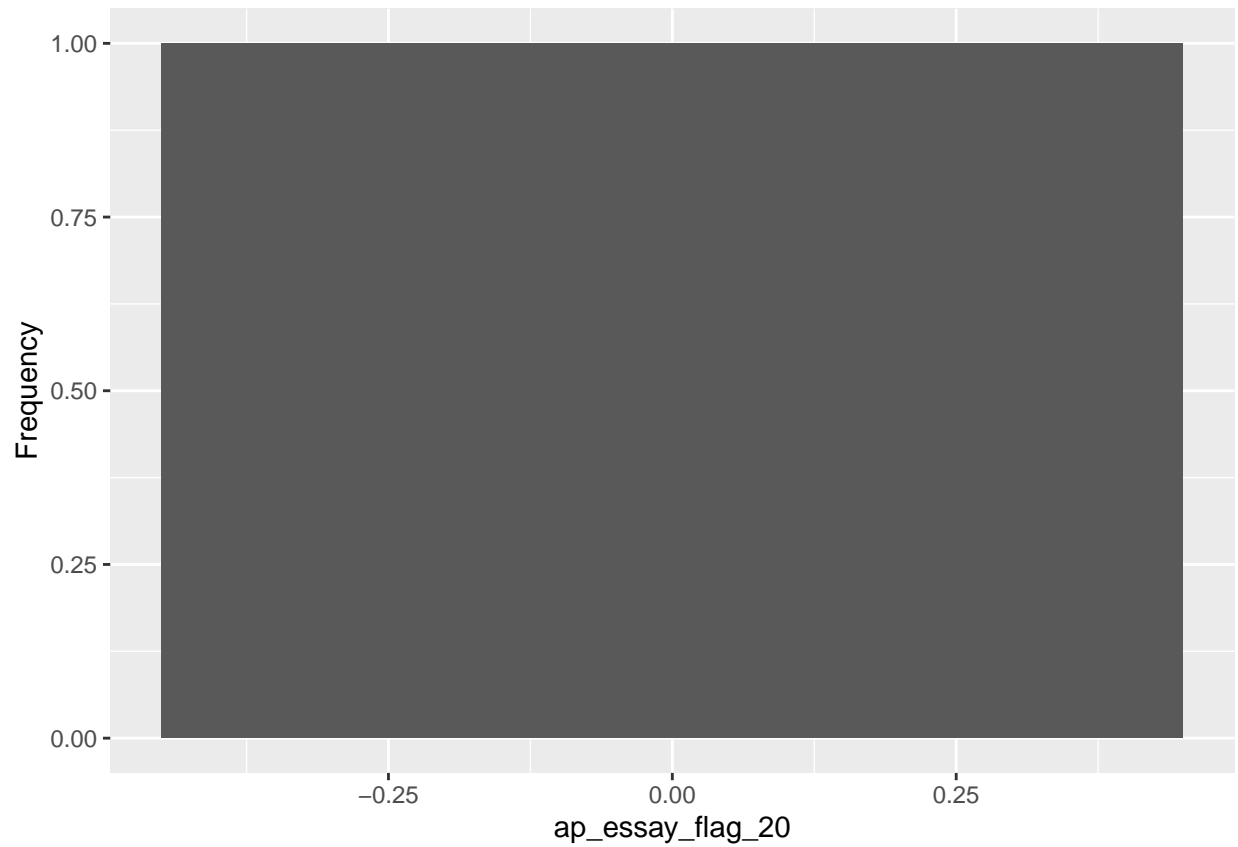
```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_year_20, type: integer, requested: yes
[1] Values (3 unique): NA, 2015, 2014
[1] Missing: 100%
```

Warning: Removed 97046 rows containing non-finite values ('stat_count()').

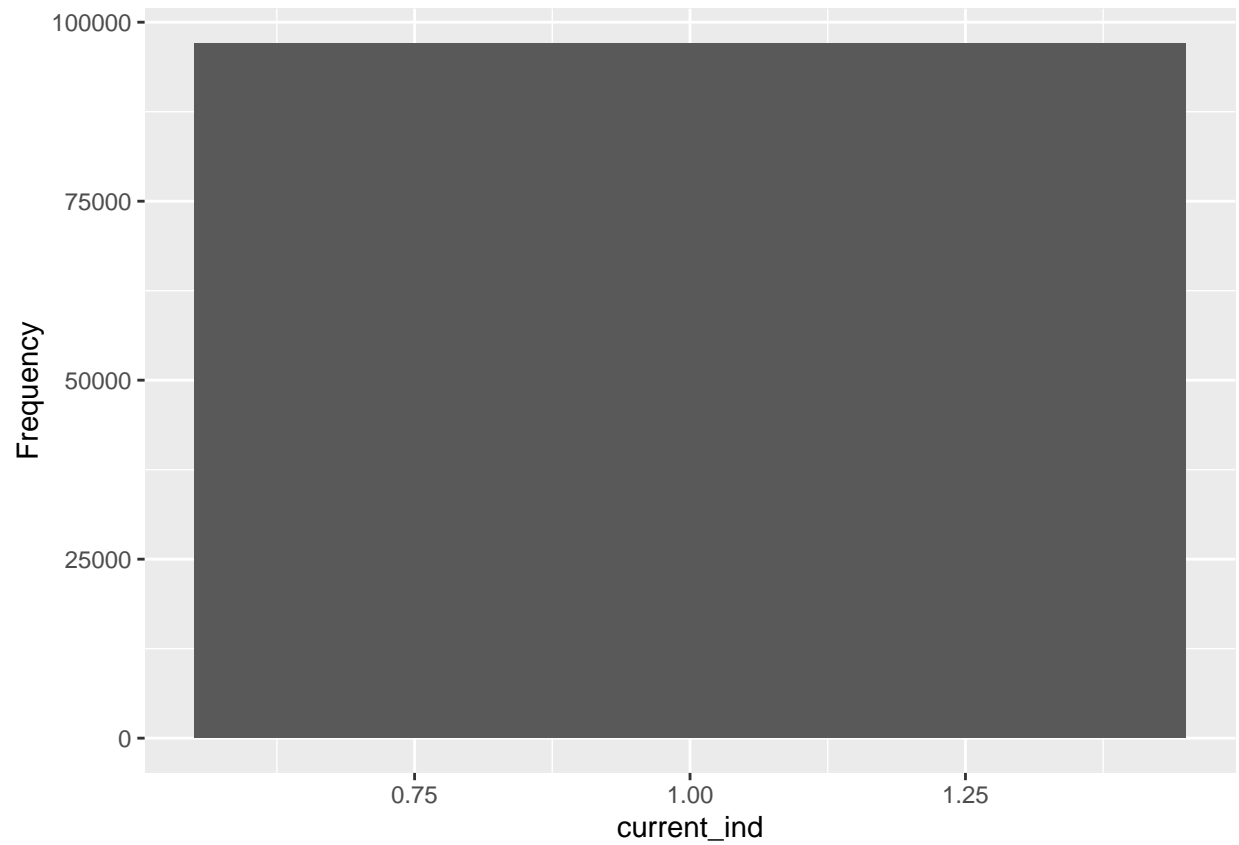


```
[1] is used in feature engineering
[1] -----
[1] Variable: ap_essay_flag_20, type: numeric, requested: yes
[1] Values (2 unique): NA, 0
[1] Missing: 100%
```

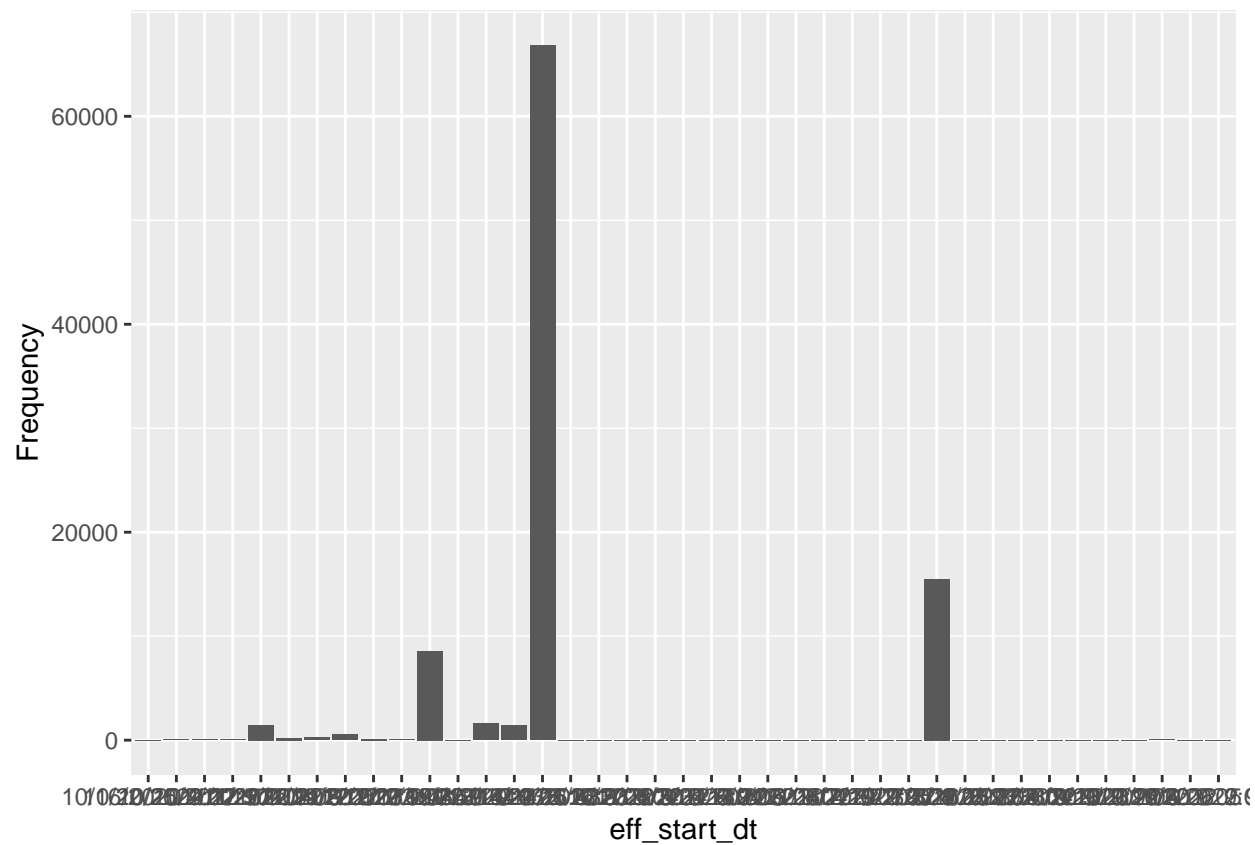
Warning: Removed 97048 rows containing non-finite values ('stat_count()').



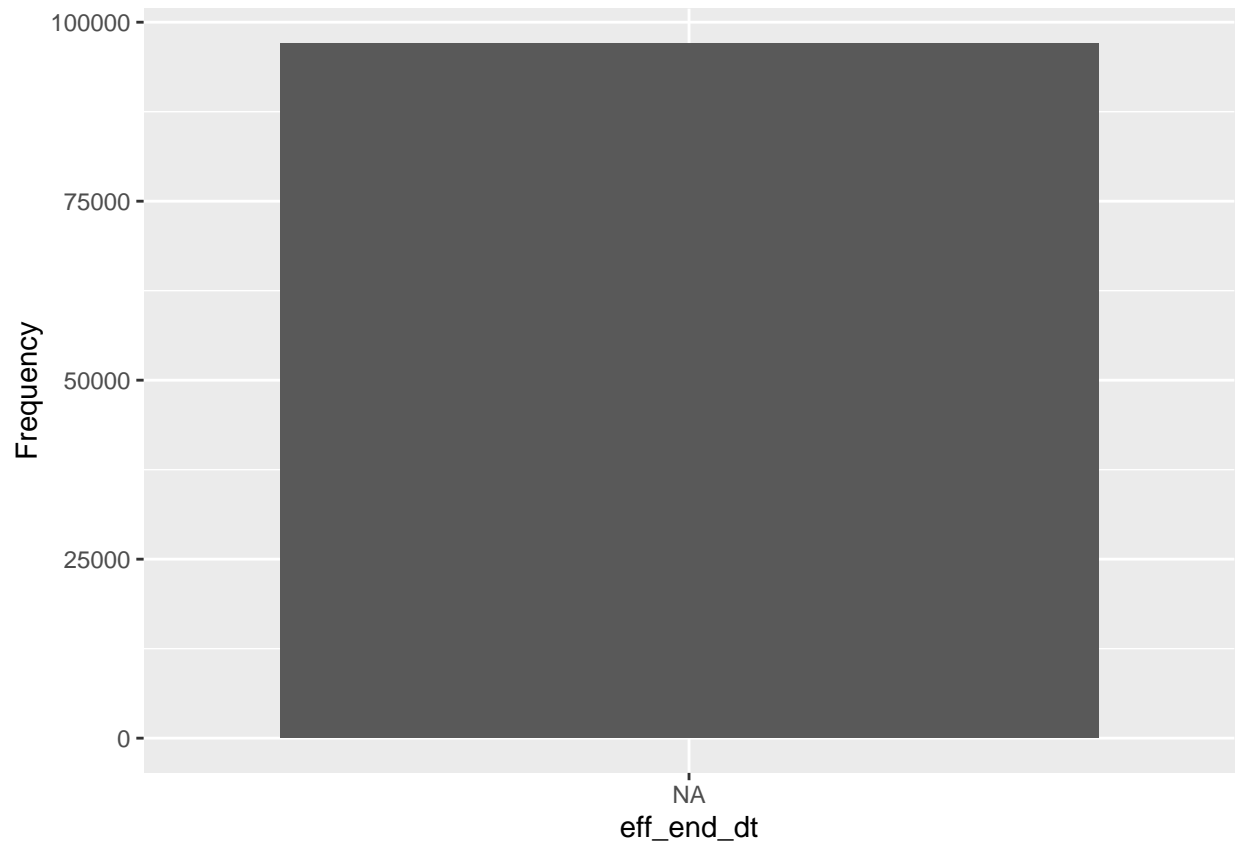
```
[1] is used in feature engineering
[1] -----
[1] Variable: current_ind, type: numeric, requested: yes
[1] Values (1 unique): 1
[1] Missing: 0%
```



```
[1] -----  
[1] Variable: eff_start_dt, type: character, requested: yes  
[1] Values (39 unique): 10/16/2020 9:02, 12/1/2021 9:59, 12/18/2021 9:03, 4/13/2022 15:43, 4/14/2022 9:  
[1] Missing: 0%
```

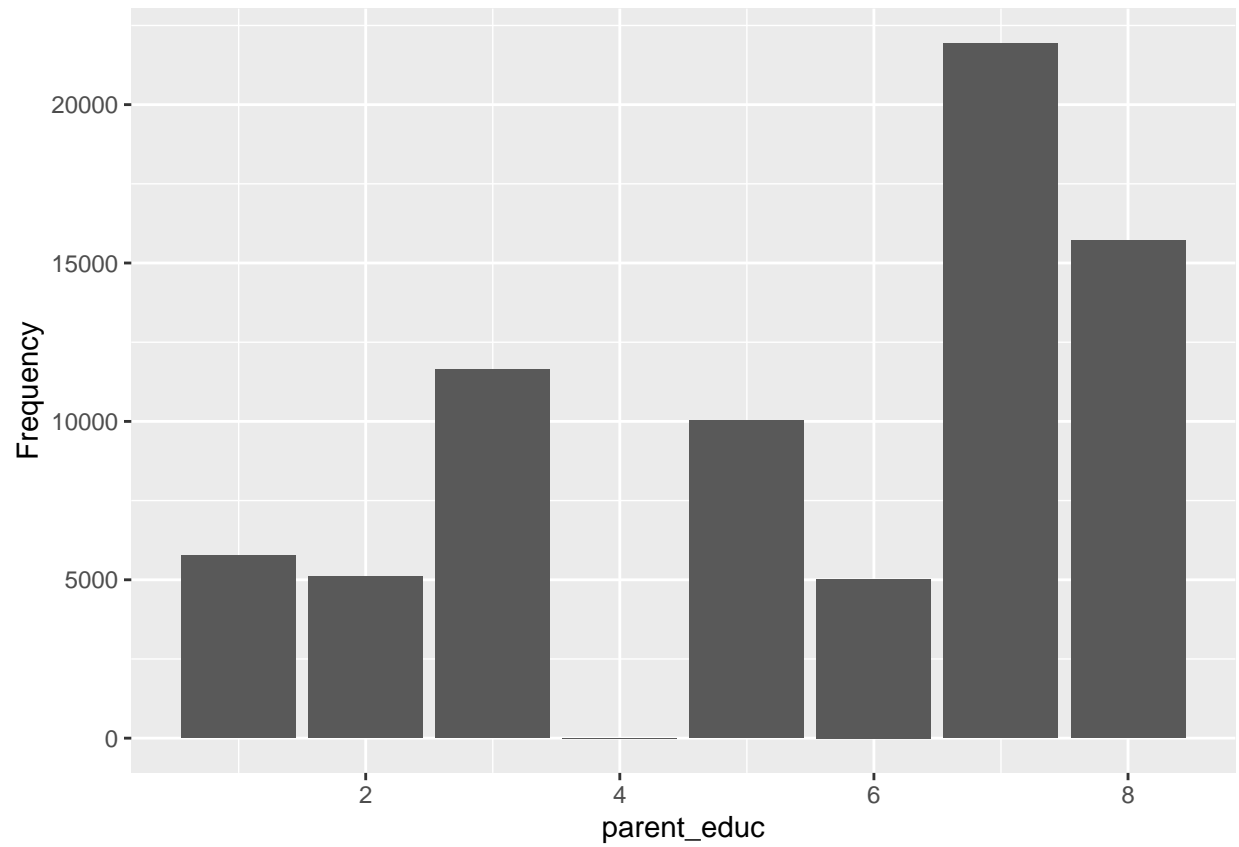


```
[1] -----
[1] Variable: eff_end_dt, type: logical, requested: yes
[1] Values (1 unique): NA
[1] Missing: 100%
```

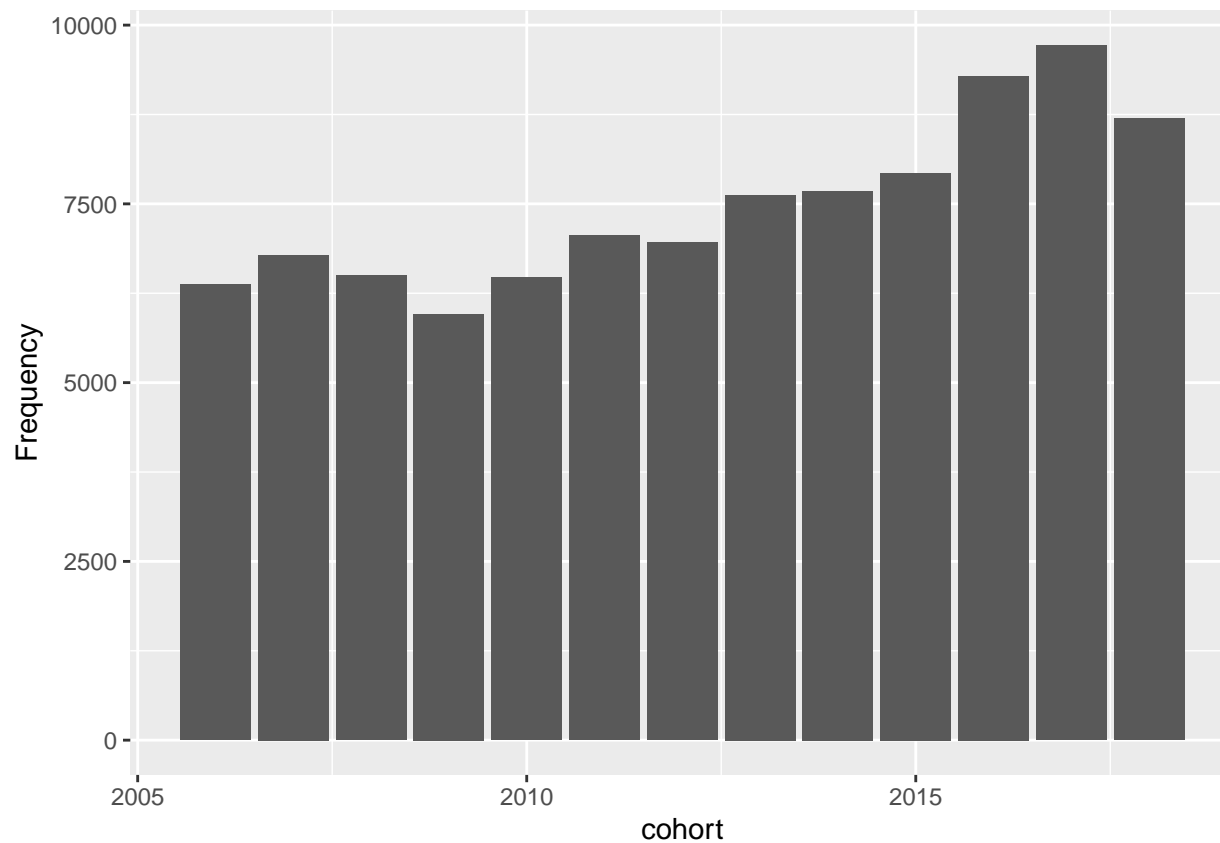


```
[1] -----
[1] Variable: parent_educ, type: numeric, requested: yes
[1] Values (9 unique): 6, NA, 8, 1, 7, ...
[1] Missing: 22.5%
  Group.1 parent_educ
1    2006  0.94713725
2    2007  0.83419307
3    2008  0.54806953
4    2009  0.42456788
5    2010  0.09210323
6    2011  0.05042493
7    2012  0.04452743
8    2013  0.05285939
9    2014  0.05924479
10   2015  0.06303580
11   2016  0.05619550
12   2017  0.04672224
13   2018  0.05244997
```

Warning: Removed 21843 rows containing non-finite values ('stat_count()').

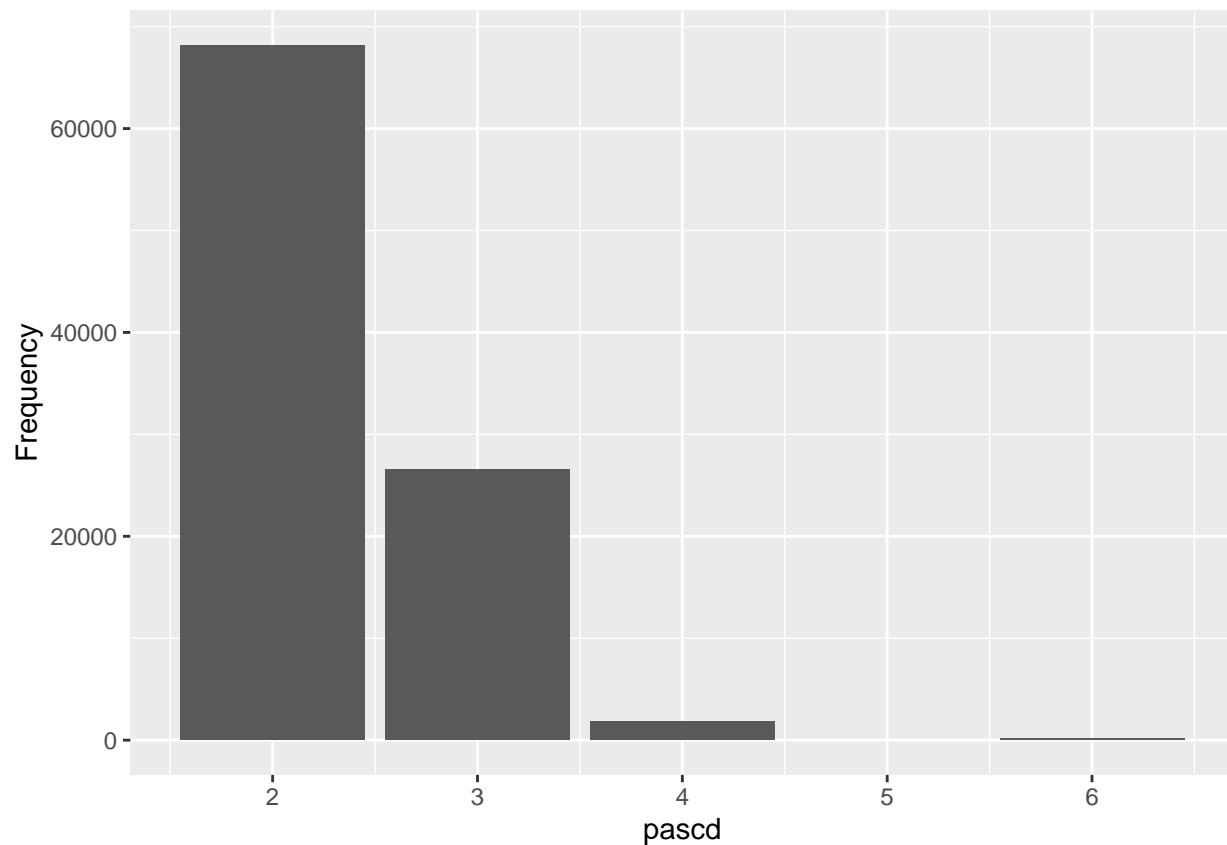


```
[1] -----  
[1] Variable: cohort, type: numeric, requested: yes  
[1] Values (13 unique): 2017, 2018, 2015, 2010, 2016, ...  
[1] Missing: 0%
```



```
[1] should not be used as a predictor
[1] -----
[1] Variable: pascd, type: numeric, requested: no
[1] Values (5 unique): 3, NA, 2, 4, 6
[1] Missing: 0.4%
```

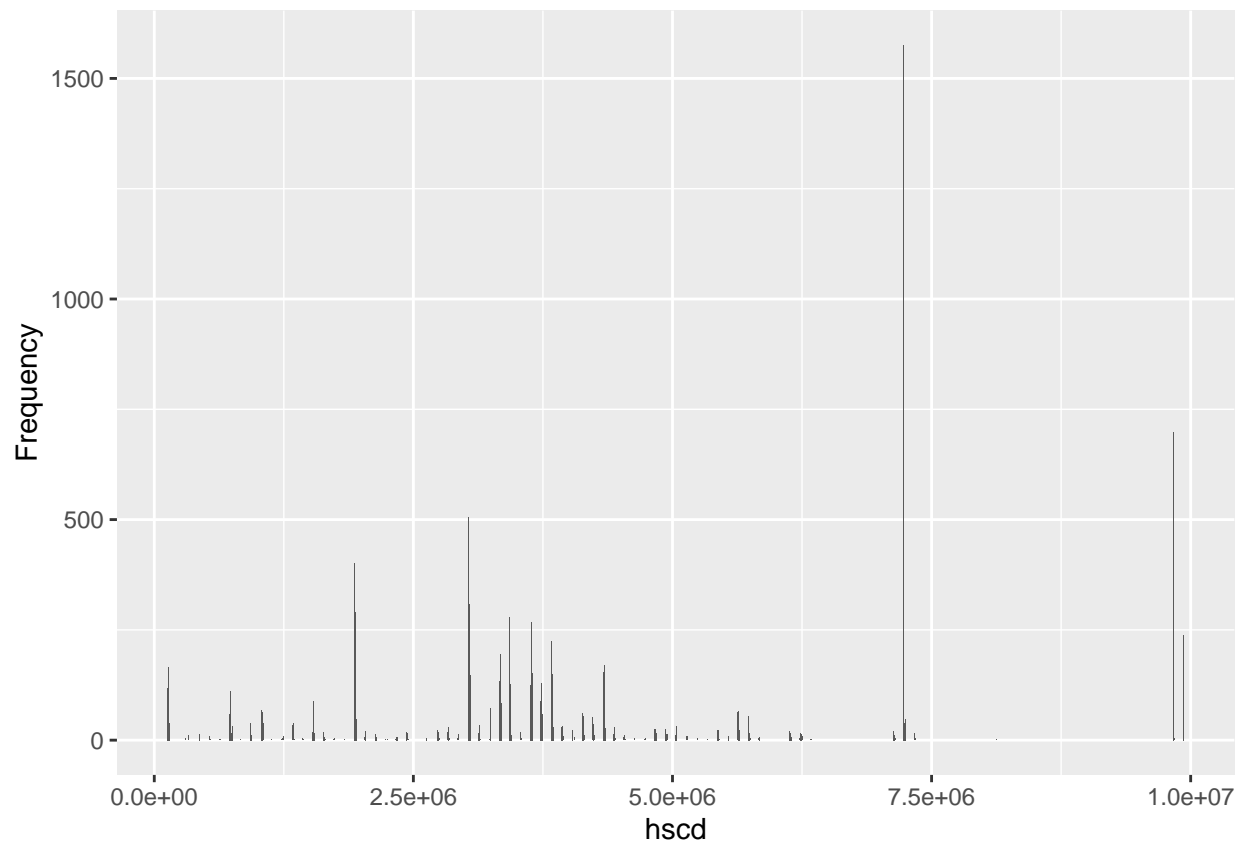
Warning: Removed 401 rows containing non-finite values ('stat_count()').



```
[1] -----
[1] Variable: hscd, type: numeric, requested: no
[1] Values (6460 unique): NA, 3330380, 4333520, 3039858, 734360, ...
[1] Missing: 33.4%
```

	Group.1	hscd
1	2006	0.9990588235
2	2007	0.9985261606
3	2008	0.9955391478
4	2009	0.9974827991
5	2010	0.9970638232
6	2011	0.0011331445
7	2012	0.0014363689
8	2013	0.0007869885
9	2014	0.0042968750
10	2015	0.0056732224
11	2016	0.0082893745
12	2017	0.0059689204
13	2018	0.0157579940

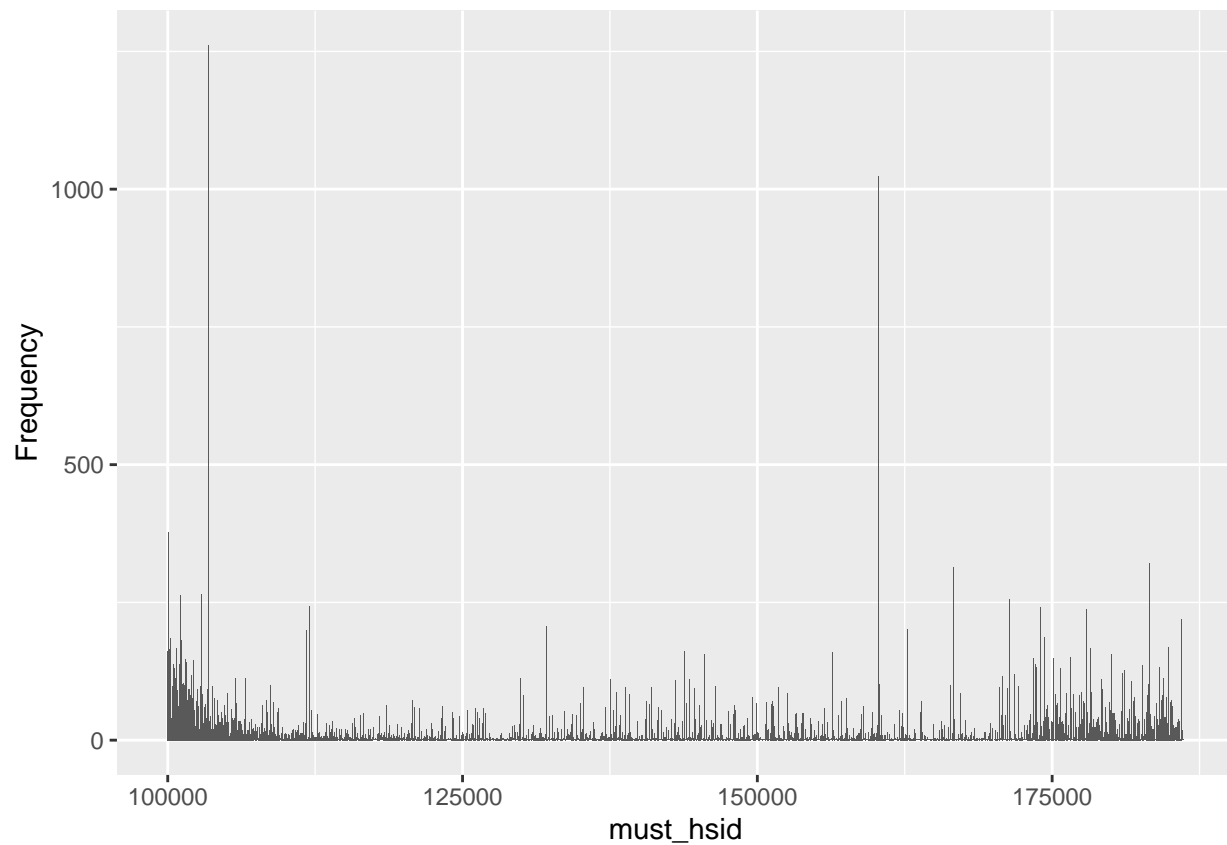
Warning: Removed 32386 rows containing non-finite values ('stat_count()').



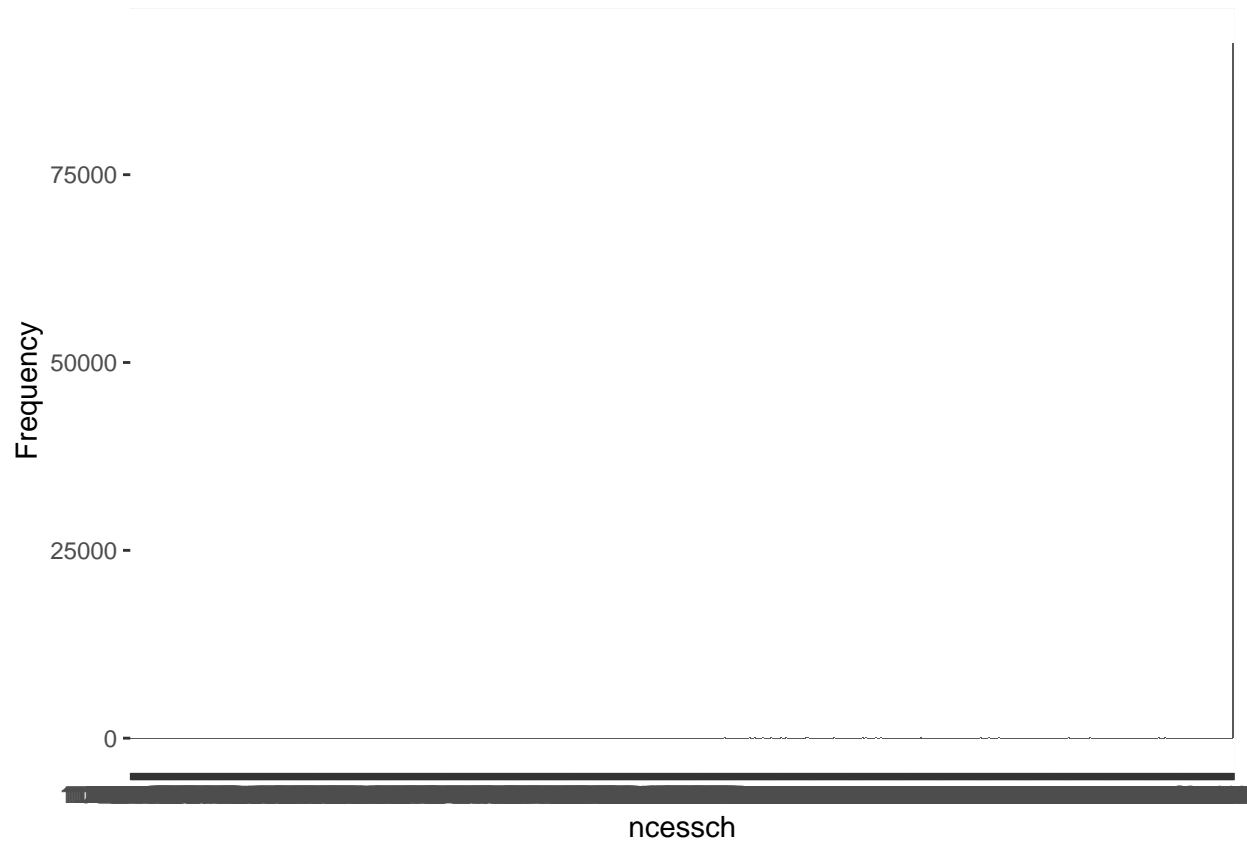
```
[1] -----
[1] Variable: must_hsid, type: numeric, requested: no
[1] Values (8205 unique): NA, 101579, 100230, 101077, 102340, ...
[1] Missing: 33.2%
```

	Group.1	must_hsid
1	2006	0.9989019608
2	2007	0.9985261606
3	2008	0.9955391478
4	2009	0.9974827991
5	2010	0.9970638232
6	2011	0.0000000000
7	2012	0.0002872738
8	2013	0.0001311647
9	2014	0.0036458333
10	2015	0.0031517902
11	2016	0.0047367854
12	2017	0.0015436863
13	2018	0.0112721417

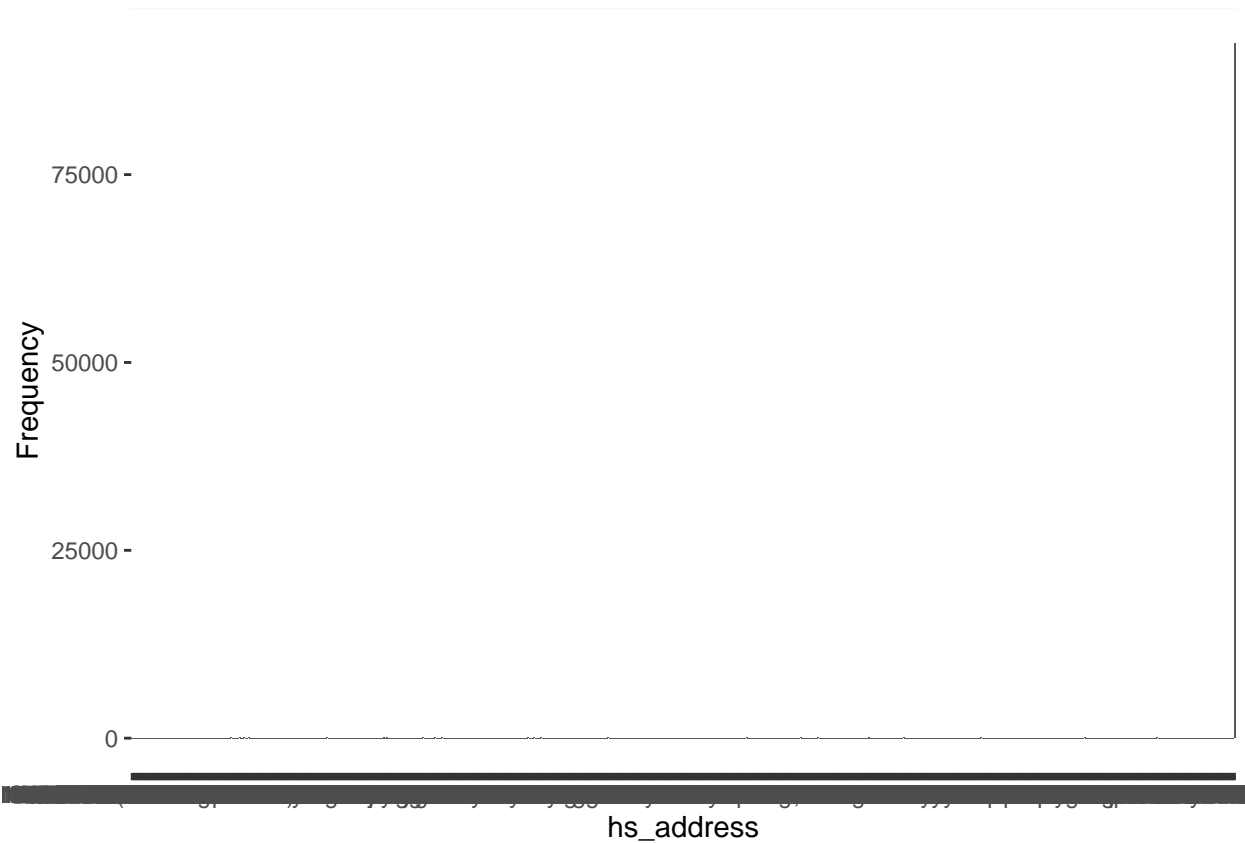
Warning: Removed 32224 rows containing non-finite values ('stat_count()').



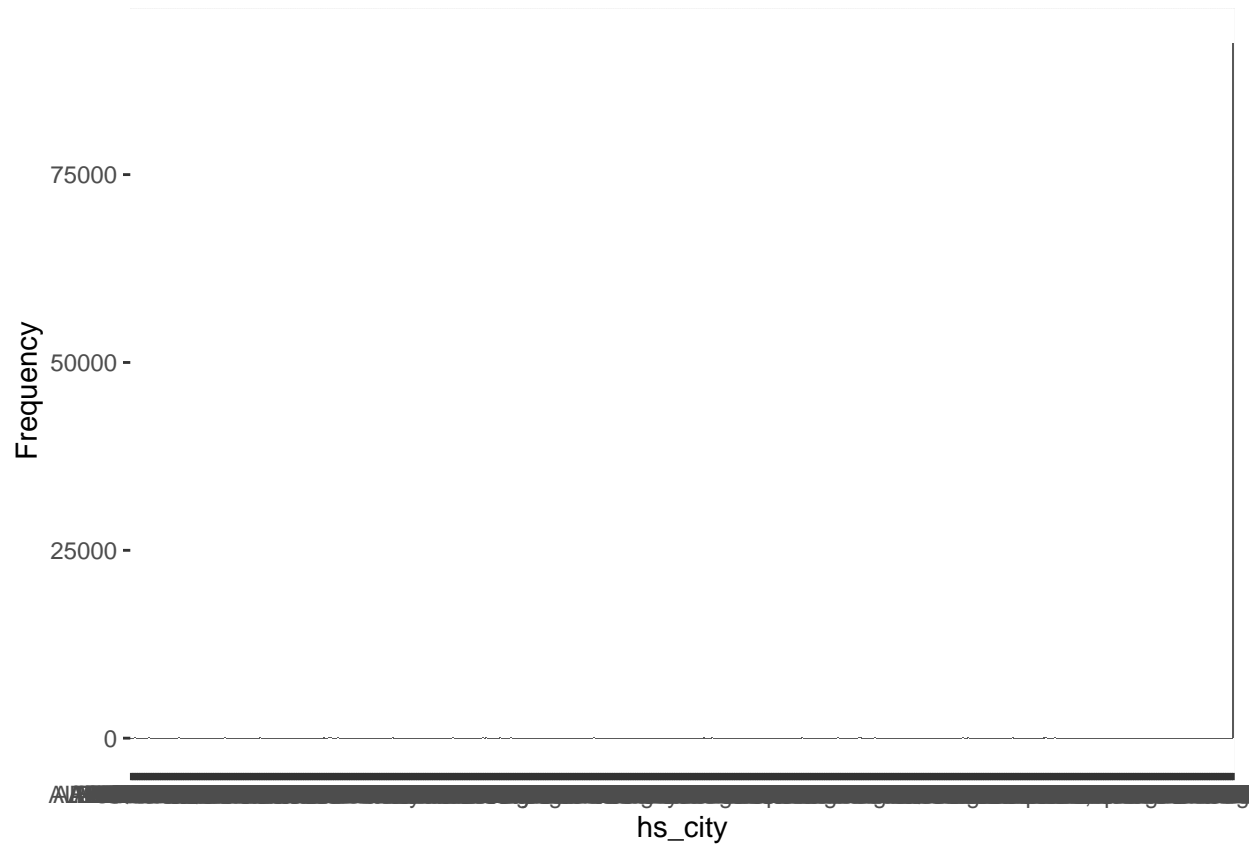
```
[1] -----
[1] Variable: ncessch, type: character, requested: no
[1] Values (664 unique): NA, 61111008819, 61212001363, 64059006709, 5.30039E+11, ...
[1] Missing: 95.3%
[1] Most missing: 2006 100%, Least missing: 2017 91.9%
```



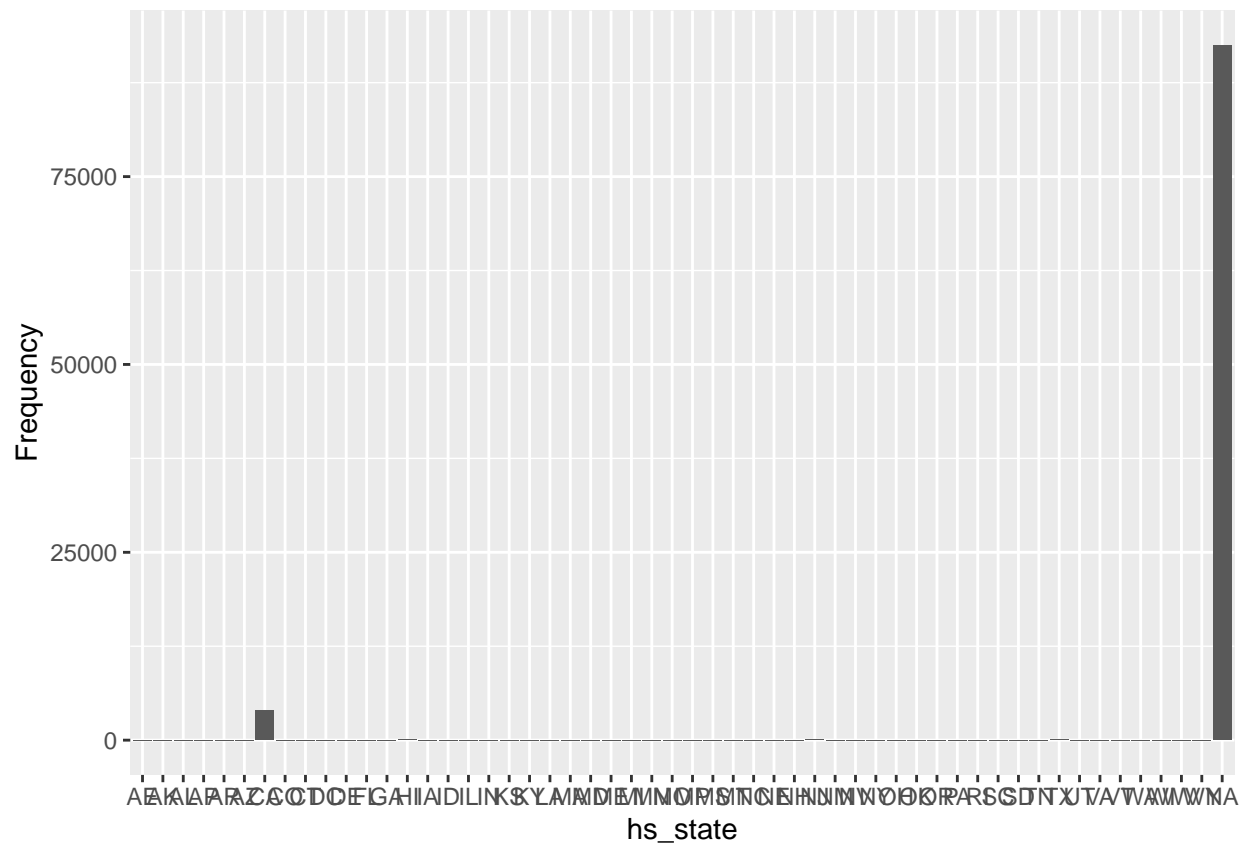
```
[1] -----
[1] Variable: hs_address, type: character, requested: no
[1] Values (758 unique): NA, 43-570 PHYLLIS JACKSON LANE, 4921 NORTH CEDAR AVENUE, 100 MONTE VISTA AVENUE
[1] Missing: 95.3%
[1] Most missing: 2006 100%, Least missing: 2017 91.9%
```



```
[1] -----
[1] Variable: hs_city, type: character, requested: no
[1] Values (625 unique): NA, PALM DESERT, EL MONTE, VACAVILLE, BELLEVUE, ...
[1] Missing: 95.3%
[1] Most missing: 2006 100%, Least missing: 2017 91.9%
```

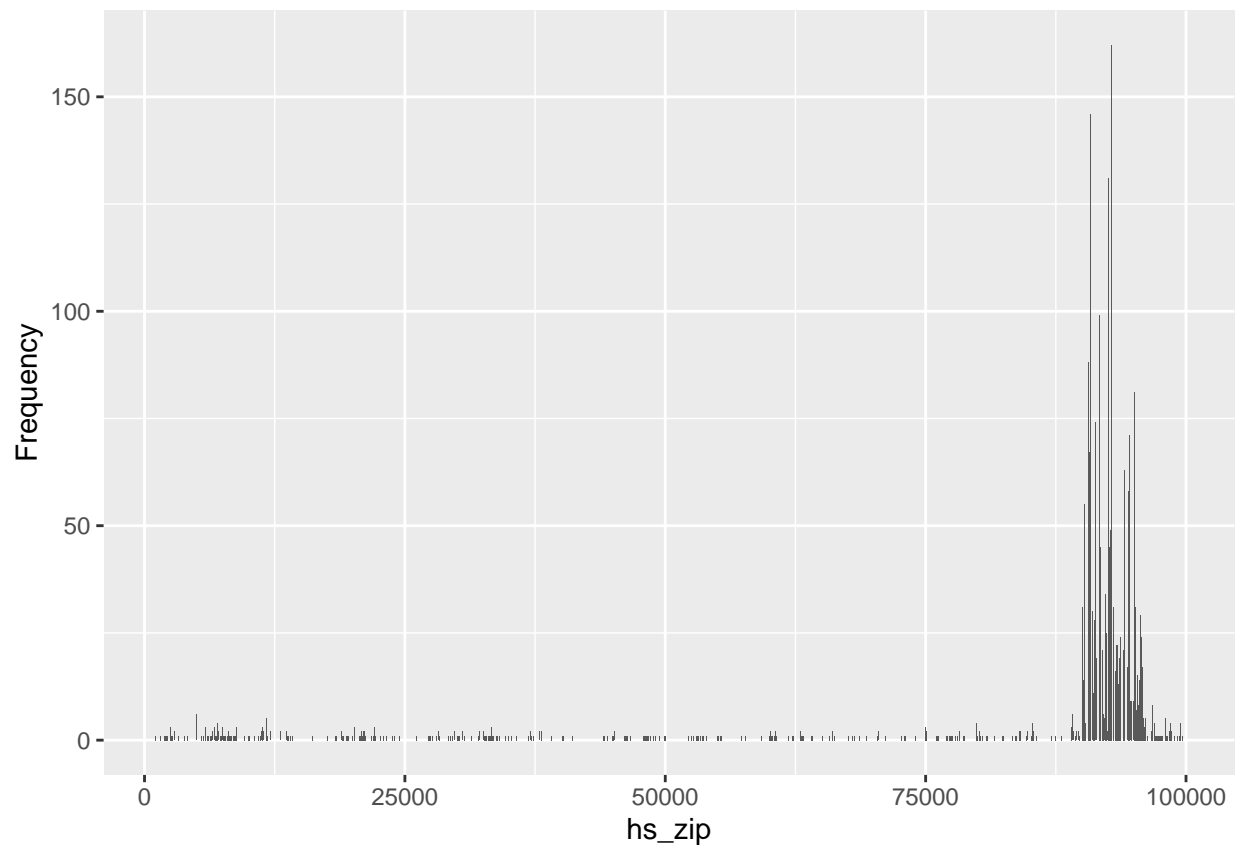


```
[1] -----
[1] Variable: hs_state, type: character, requested: no
[1] Values (54 unique): NA, CA, WA, CO, HI, ...
[1] Missing: 95.3%
[1] Most missing: 2006 100%, Least missing: 2017 91.9%
```



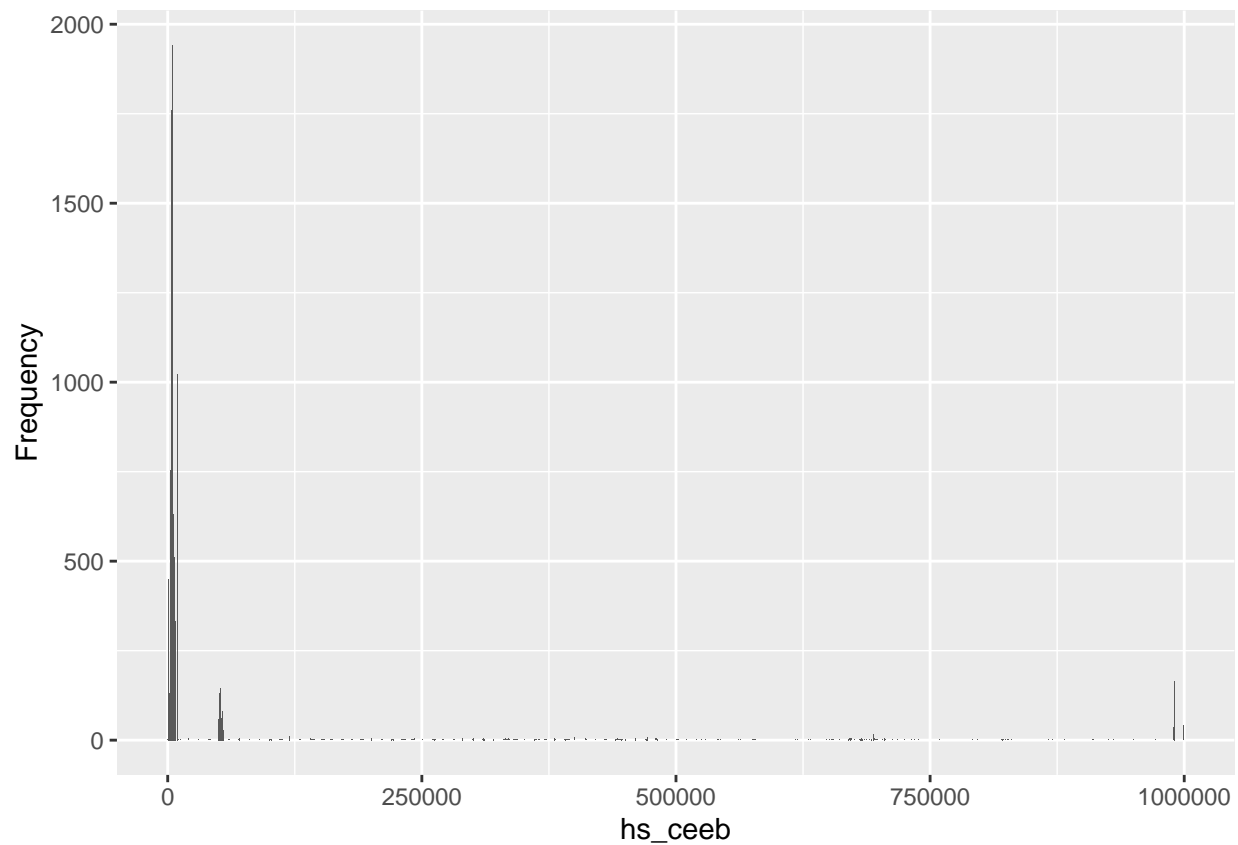
```
[1] -----
[1] Variable: hs_zip, type: numeric, requested: no
[1] Values (721 unique): NA, 92260, 91732, 95688, 98004, ...
[1] Missing: 95.3%
[1] Most missing: 2006 100%, Least missing: 2017 91.9%
```

Warning: Removed 92471 rows containing non-finite values ('stat_count()').

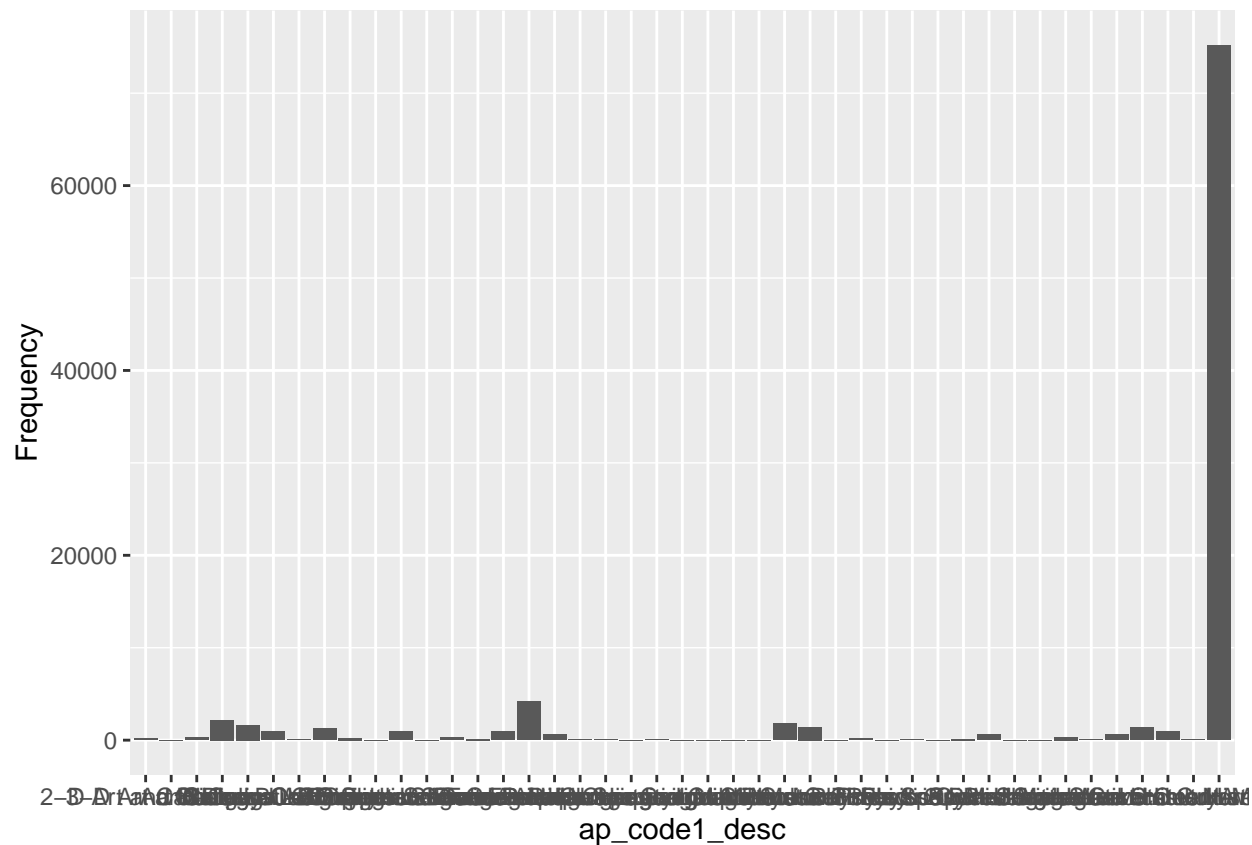


```
[1] -----
[1] Variable: hs_ceeb, type: numeric, requested: no
[1] Values (2407 unique): NA, 52319, 4860, 4334, 2711, ...
[1] Missing: 37%
  Group.1    hs_ceeb
1      2006 0.99890196
2      2007 0.99852616
3      2008 0.99553915
4      2009 0.99765061
5      2010 0.99737289
6      2011 0.02252125
7      2012 0.04079288
8      2013 0.05194124
9      2014 0.05976563
10     2015 0.07249117
11     2016 0.08590806
12     2017 0.06946588
13     2018 0.06579250
```

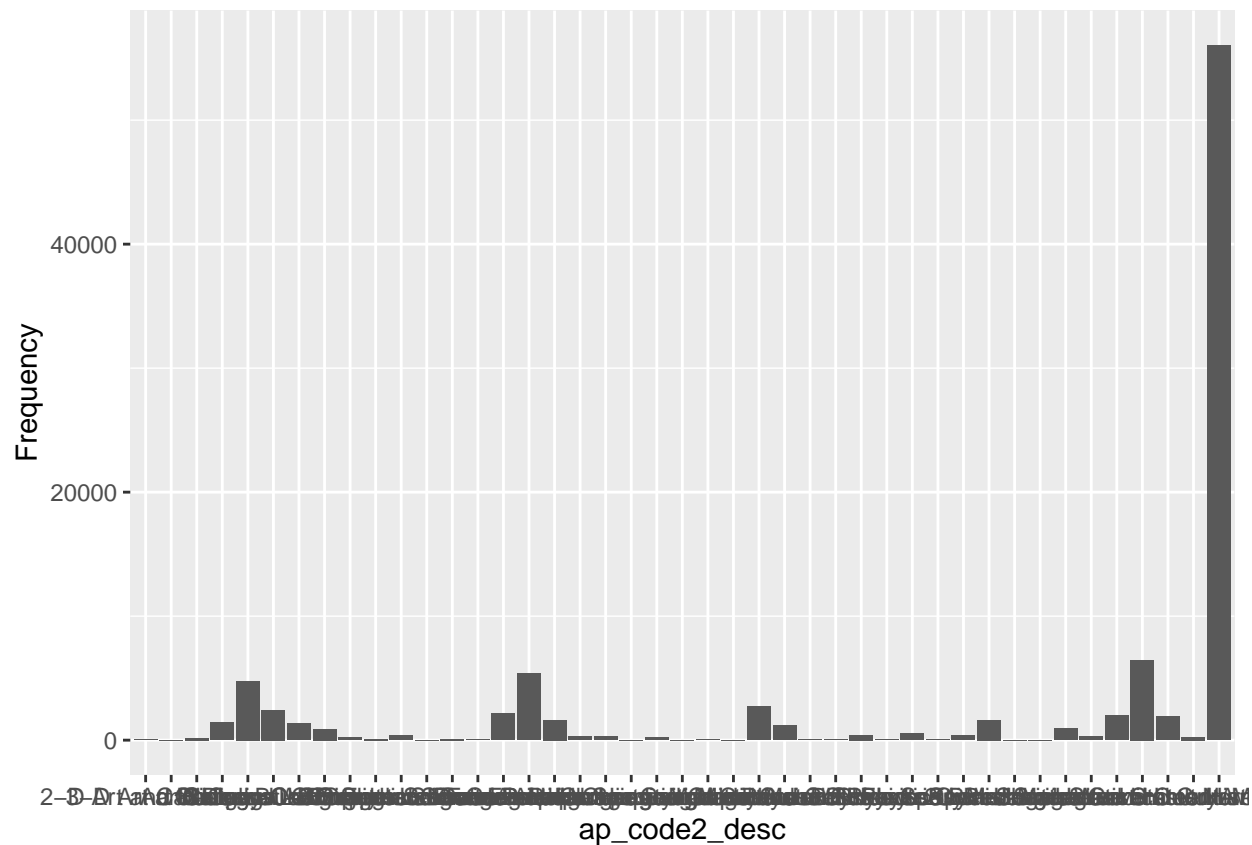
Warning: Removed 35932 rows containing non-finite values ('stat_count()').



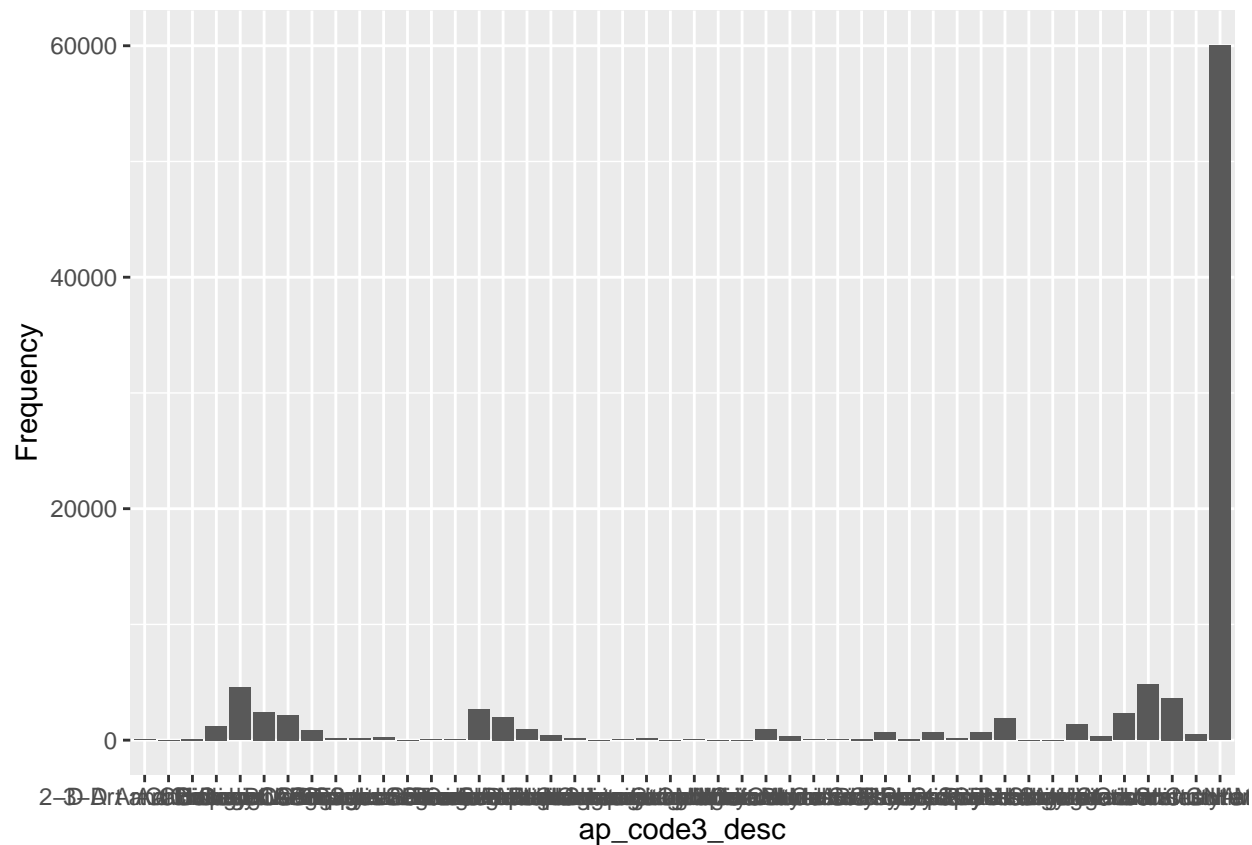
```
[1] -----
[1] Variable: ap_code1_desc, type: character, requested: no
[1] Values (43 unique): NA, Calculus AB, Microeconomics, Macroeconomics, Biology, ...
[1] Missing: 77.5%
  Group.1 ap_code1_desc
1    2006      1.0000000
2    2007      0.9998526
3    2008      1.0000000
4    2009      1.0000000
5    2010      1.0000000
6    2011      0.9923513
7    2012      0.9918127
8    2013      0.9815058
9    2014      0.9085937
10   2015      0.5260968
11   2016      0.4491334
12   2017      0.3627663
13   2018      0.3270071
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code2_desc, type: character, requested: no
[1] Values (43 unique): NA, Physics C: Mechanics, United States Government and Politics, Physics 1, Eng
[1] Missing: 57.7%
  Group.1 ap_code2_desc
1      2006      0.9995294
2      2007      0.9988209
3      2008      0.9972312
4      2009      0.9981541
5      2010      0.9983001
6      2011      0.3270538
7      2012      0.3761850
8      2013      0.3610965
9      2014      0.3757813
10     2015      0.3765759
11     2016      0.3837873
12     2017      0.3660595
13     2018      0.3814125
```

```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code3_desc, type: character, requested: no
[1] Values (46 unique): NA, Physics C: Mechanics, Statistics, Calculus AB, Chemistry, ...
[1] Missing: 61.9%
Group.1 ap_code3_desc
1      2006      0.9995294
2      2007      0.9988209
3      2008      0.9976927
4      2009      0.9981541
5      2010      0.9987637
6      2011      0.4025496
7      2012      0.4478598
8      2013      0.4252361
9      2014      0.4291667
10     2015      0.4358296
11     2016      0.4385833
12     2017      0.4246167
13     2018      0.4409938
```



[1] should not be used as a predictor

[1] is used in feature engineering

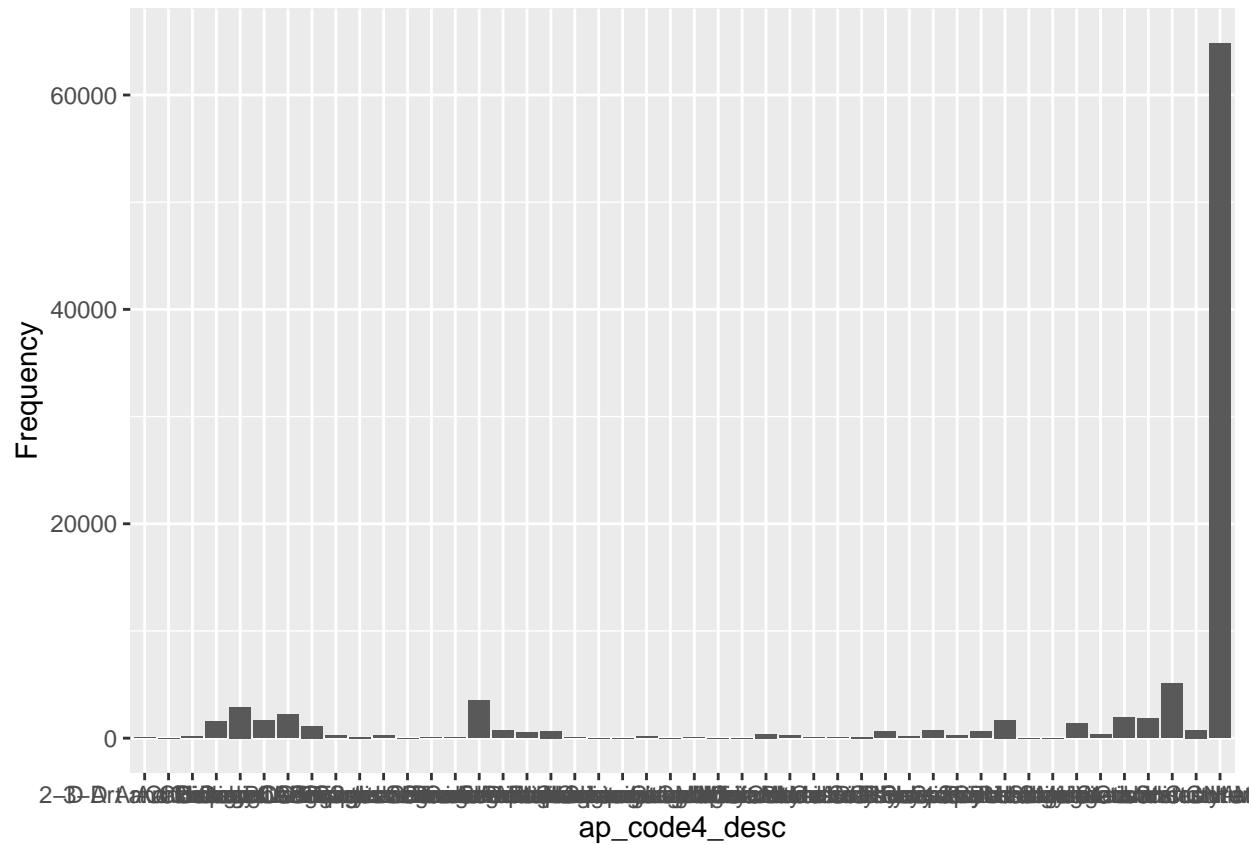
[1] -----

[1] Variable: ap_code4_desc, type: character, requested: no

[1] Values (46 unique): NA, Calculus BC, English Language and Composition, Spanish Literature and Culture

[1] Missing: 66.8%

Group.1	ap_code4_desc
1	2006 0.9995294
2	2007 0.9988209
3	2008 0.9984618
4	2009 0.9983219
5	2010 0.9989183
6	2011 0.4864023
7	2012 0.5277219
8	2013 0.5019675
9	2014 0.4968750
10	2015 0.5065557
11	2016 0.5073743
12	2017 0.4943913
13	2018 0.5152979



[1] should not be used as a predictor

[1] is used in feature engineering

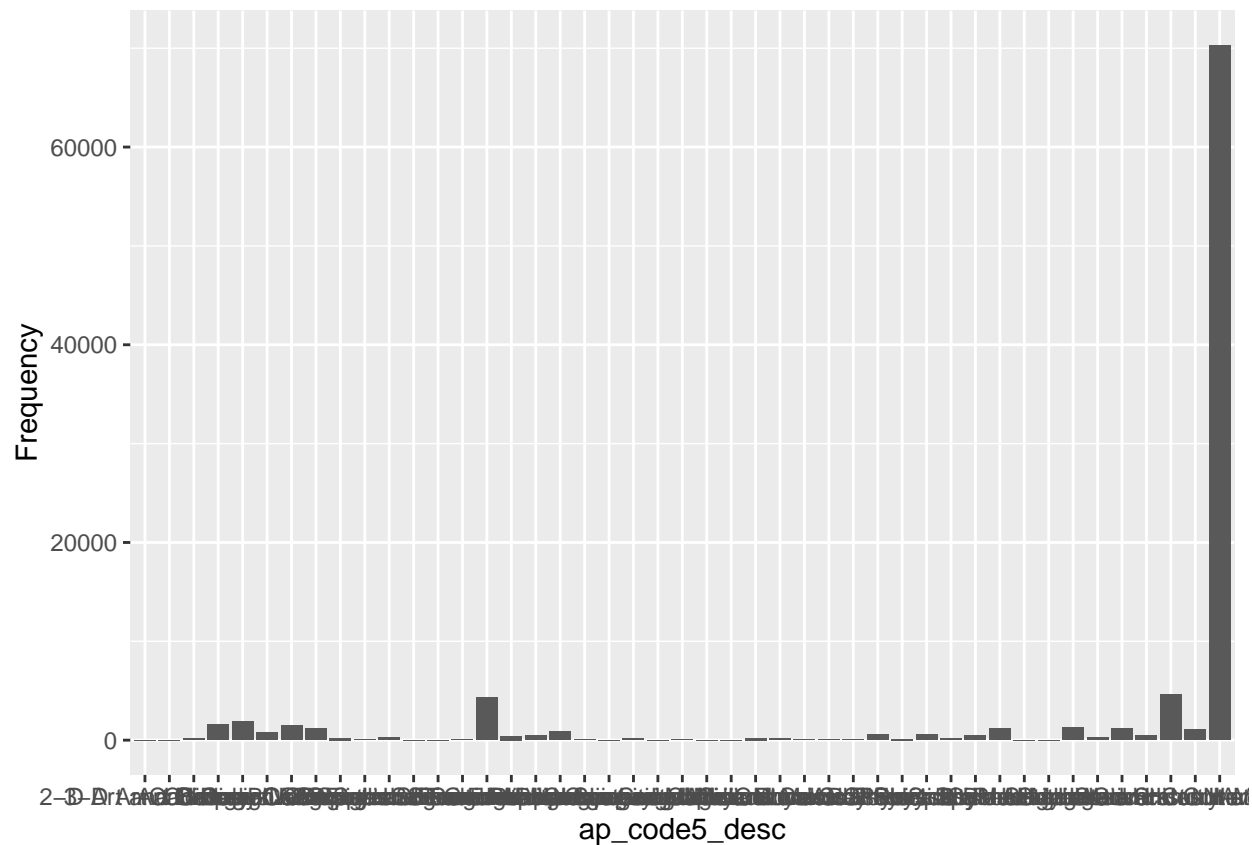
[1] _____

```
[1] Variable: ap_code5_desc, type: character, requested: no
```

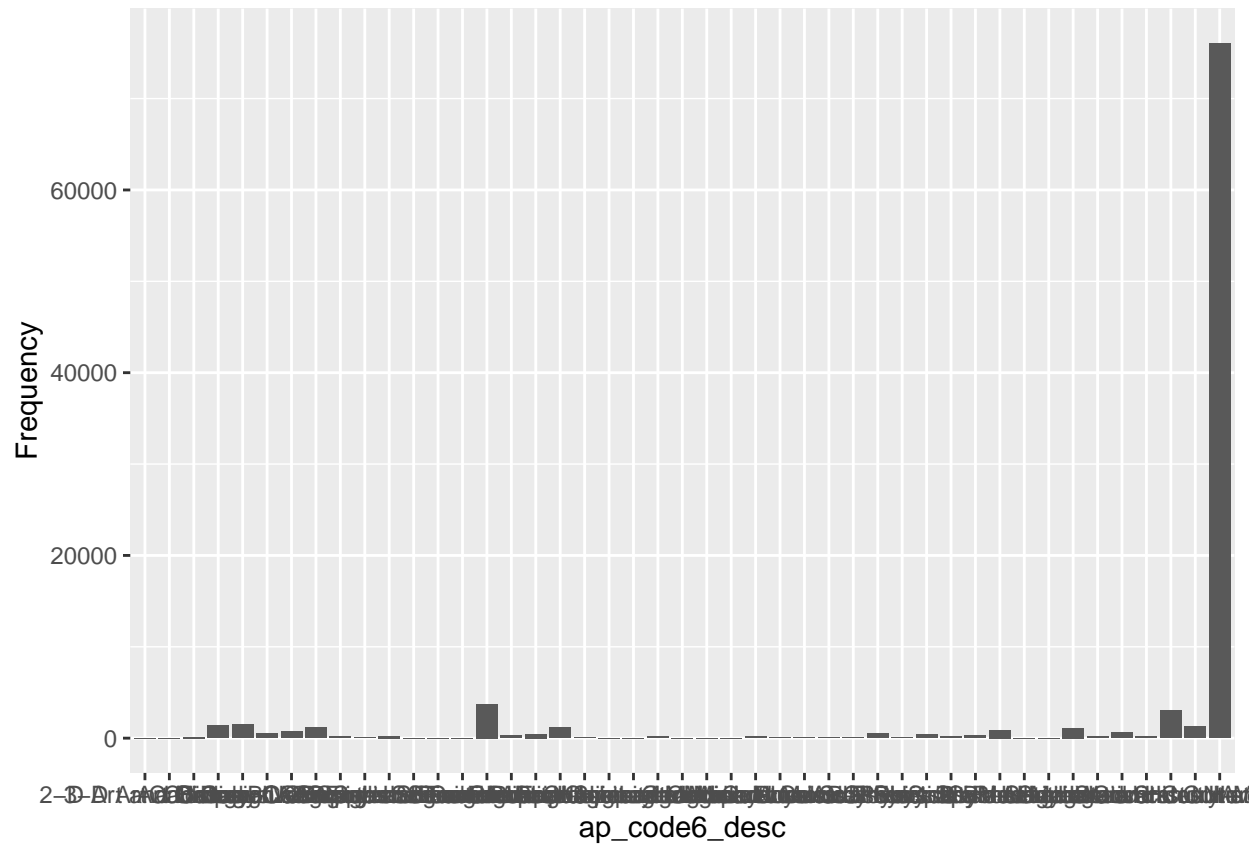
```
[1] Values (45 unique): NA, United States History, Calculus BC, United States Government and Politics, ...
```

```
[1] Missing: 72.4%
```

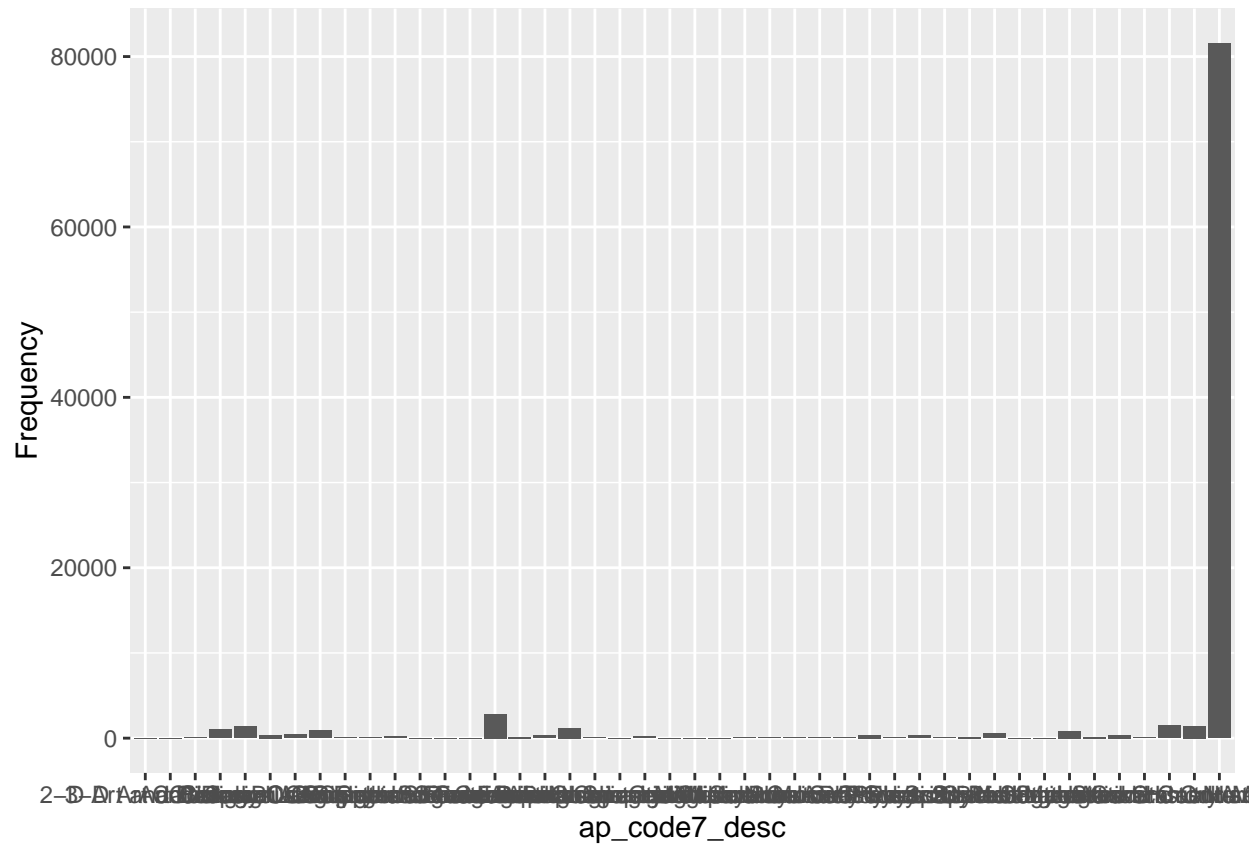
```
[1] Most missing: 2006 100%, Least missing: 2017 57.5%
```



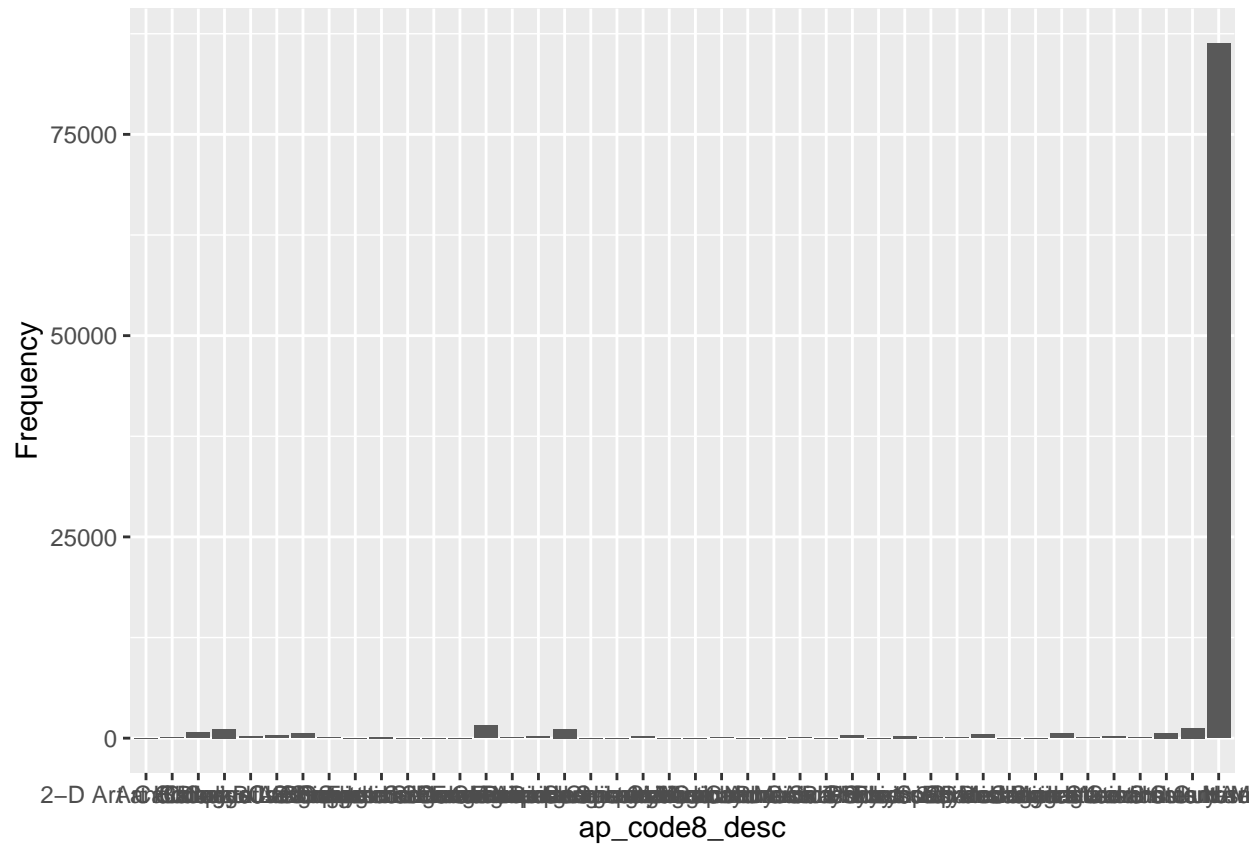
```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code6_desc, type: character, requested: no
[1] Values (45 unique): NA, English Language and Composition, Calculus BC: AB Subscore, Chemistry, Biol
[1] Missing: 78.4%
[1] Most missing: 2007 100%, Least missing: 2017 65.5%
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code7_desc, type: character, requested: no
[1] Values (44 unique): NA, Calculus AB, English Language and Composition, United States History, Chemi
[1] Missing: 84%
[1] Most missing: 2007 100%, Least missing: 2017 73.9%
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code8_desc, type: character, requested: no
[1] Values (42 unique): NA, World History: Modern, European History, Calculus AB, Spanish Language and C
[1] Missing: 88.9%
[1] Most missing: 2006 100%, Least missing: 2018 80.8%
```



[1] should not be used as a predictor

[1] is used in feature engineering

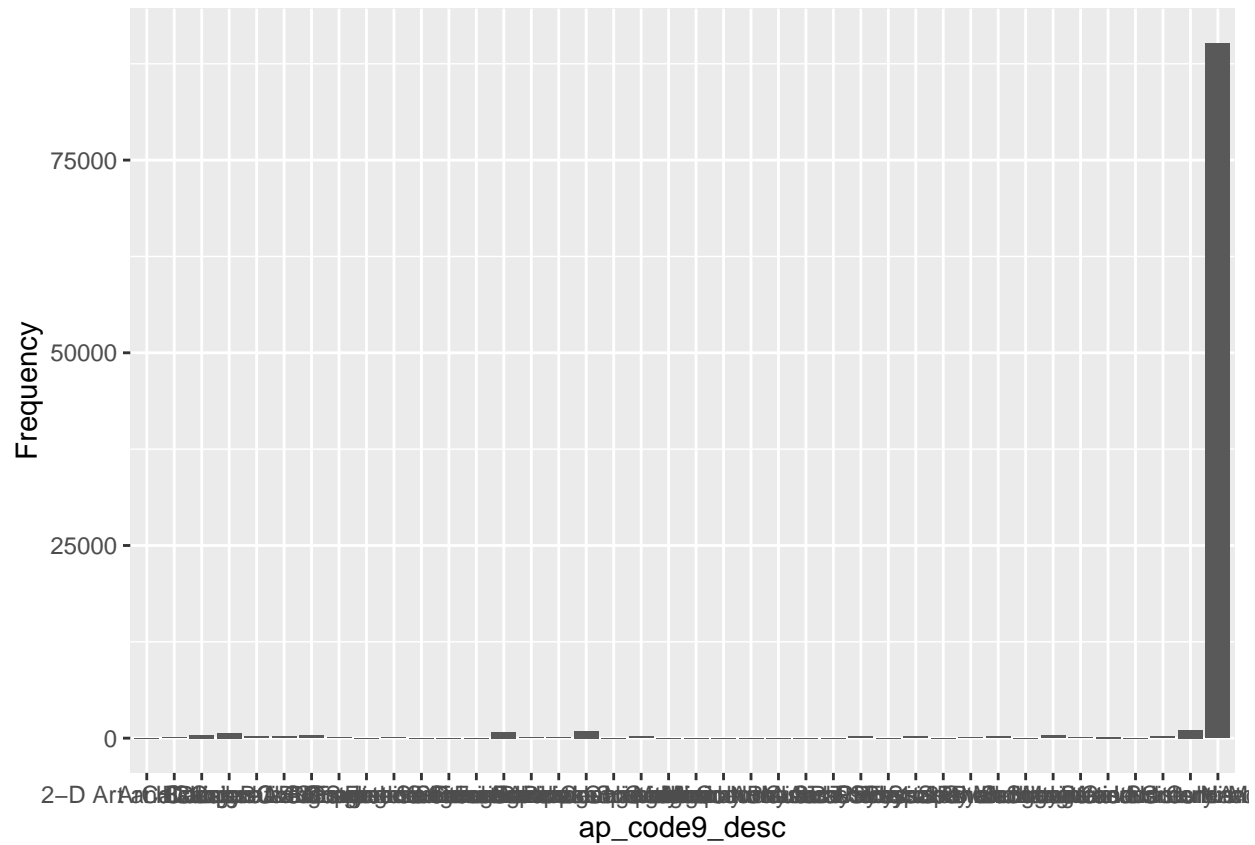
[1] _____

```
[1] Variable: ap_code9_desc, type: character, requested: no
```

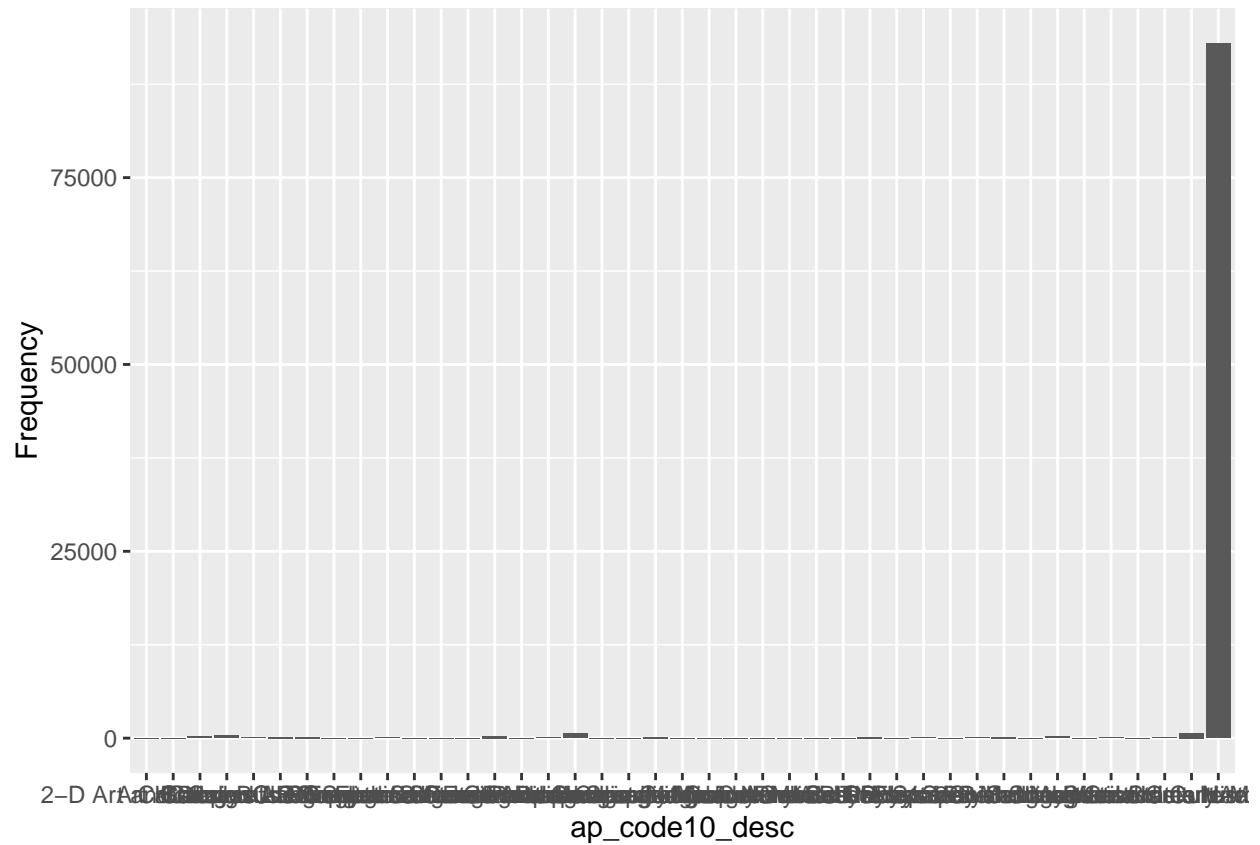
[1] Values (40 unique): NA, Human Geography, European History, World History: Modern, English Language & Literature

```
[1] Missing: 92.9%
```

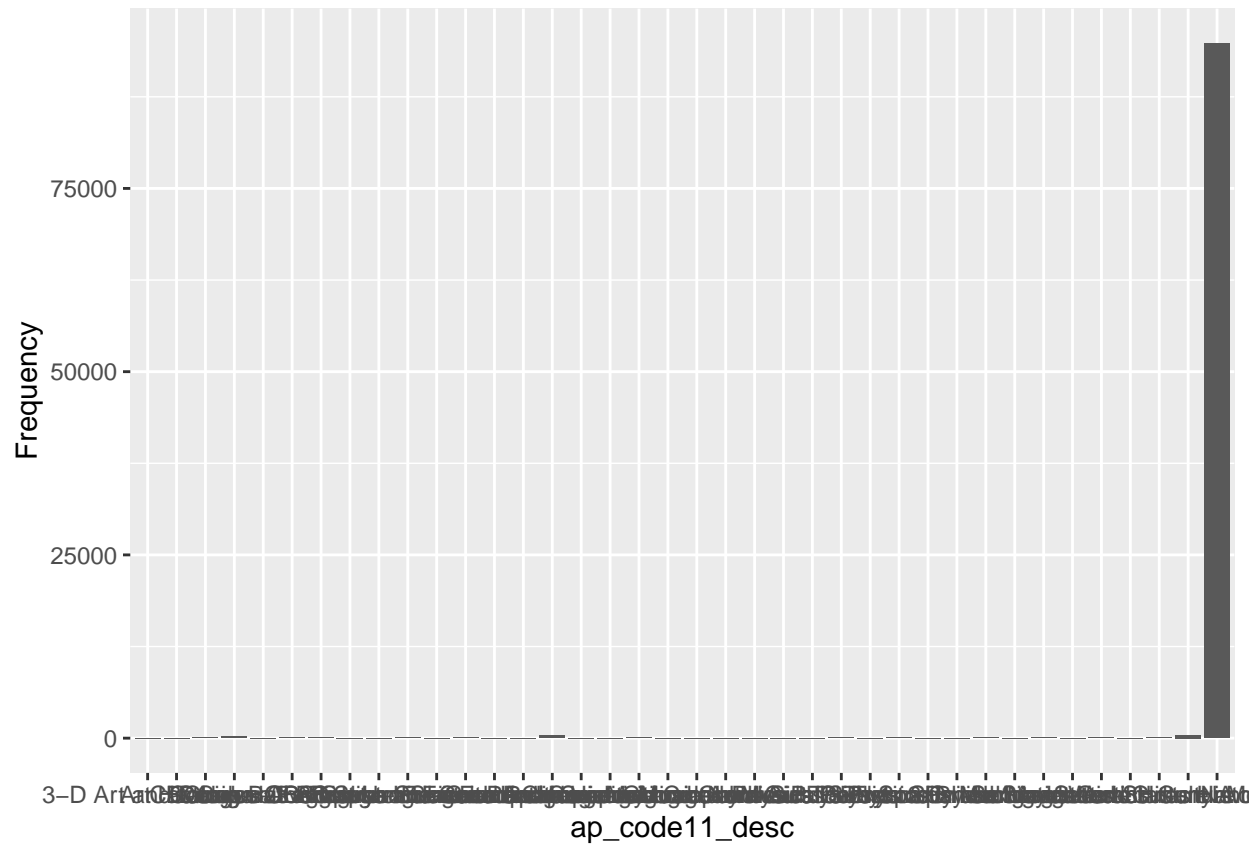
```
[1] Most missing: 2006 100%, Least missing: 2018 86.9%
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code10_desc, type: character, requested: no
[1] Values (41 unique): NA, Spanish Language and Culture, Physics 1, Chemistry, Environmental Science,
[1] Missing: 95.8%
[1] Most missing: 2006 100%, Least missing: 2018 91.6%
```

```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code11_desc, type: character, requested: no
[1] Values (38 unique): NA, Human Geography, European History, World History: Modern, Physics 1, ...
[1] Missing: 97.7%
[1] Most missing: 2006 100%, Least missing: 2018 94.7%
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code12_desc, type: character, requested: no
[1] Values (36 unique): NA, Human Geography, Psychology, European History, World History: Modern, ...
[1] Missing: 98.8%
[1] Most missing: 2006 100%, Least missing: 2018 97%
```

[1] is used in feature engineering

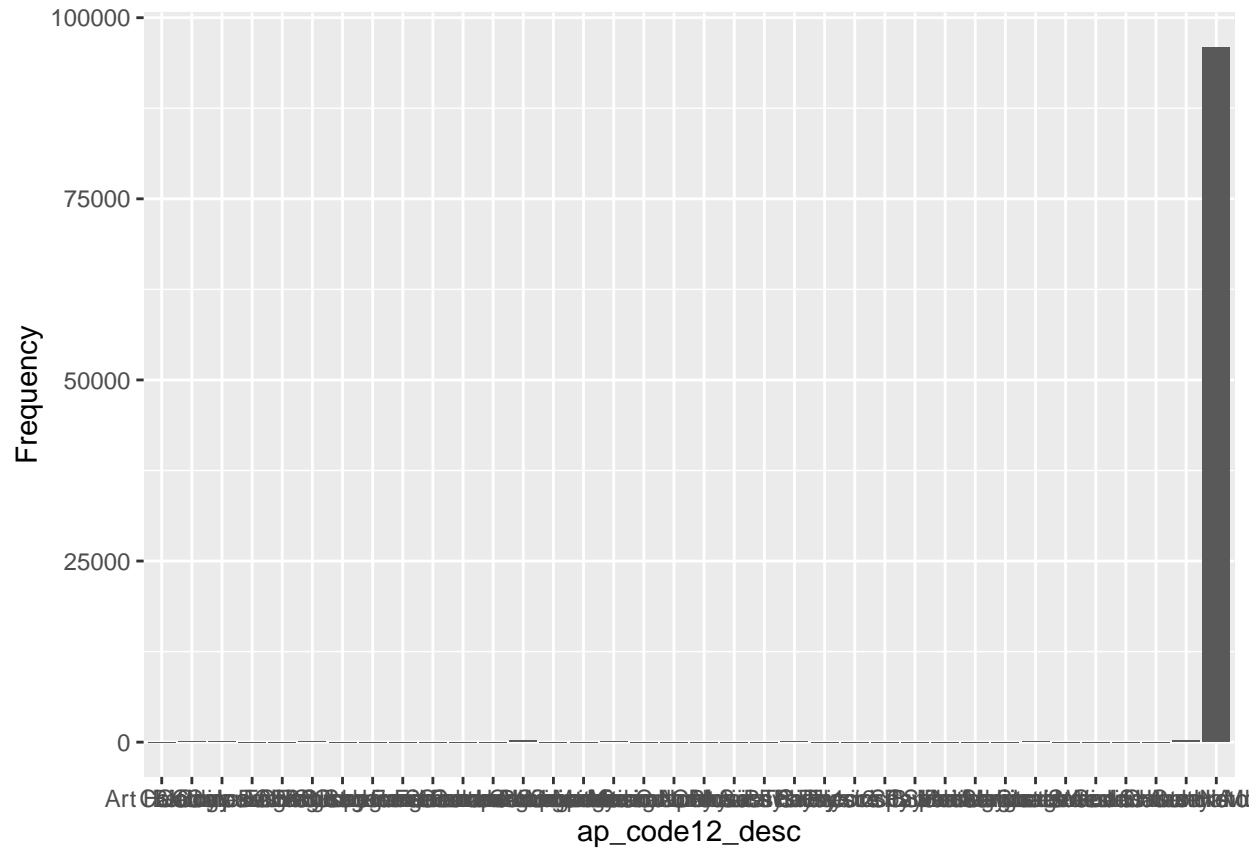
[1] -----

```
[1] Variable: ap_code12_desc, type: character, requested: no
```

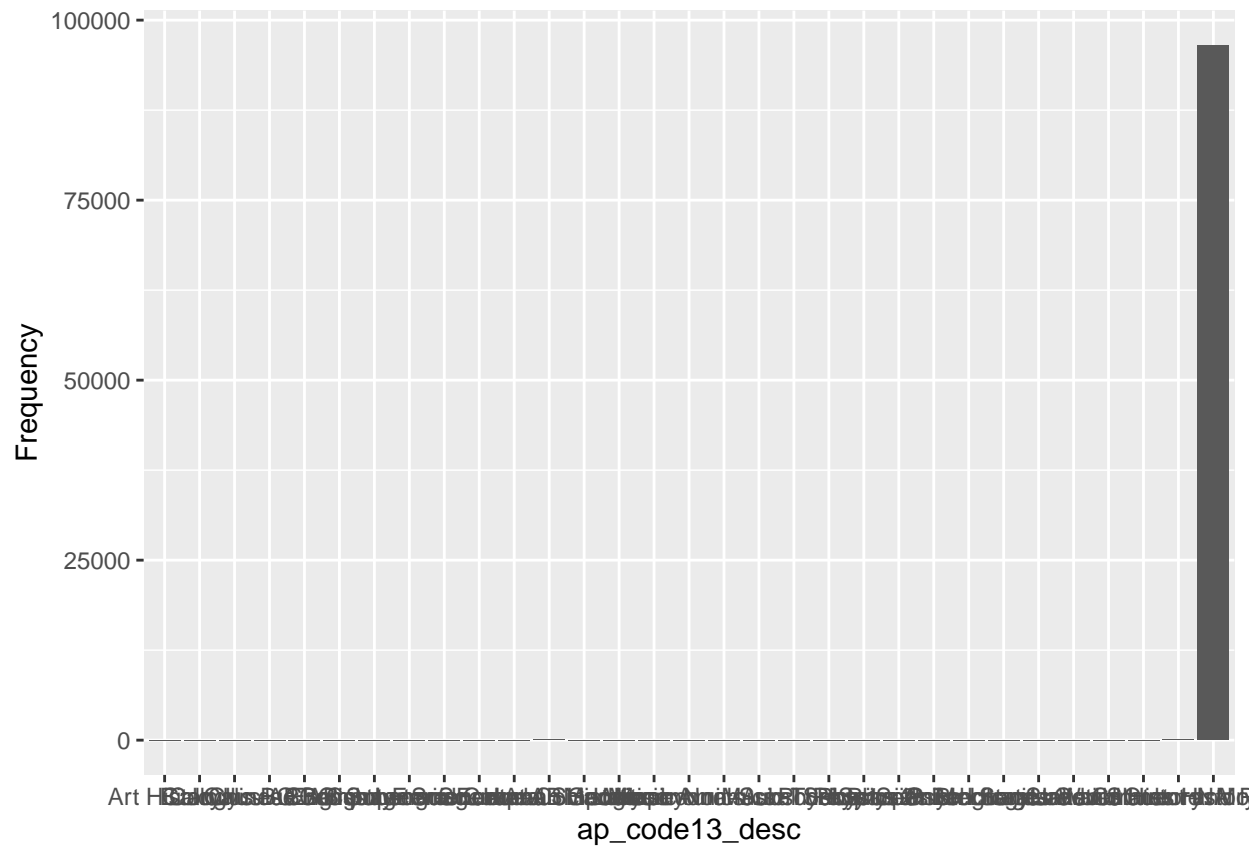
```
[1] Values (36 unique): NA, Human Geography, Psychology, European History, World History: Modern, ...
```

```
[1] Missing: 98.8%
```

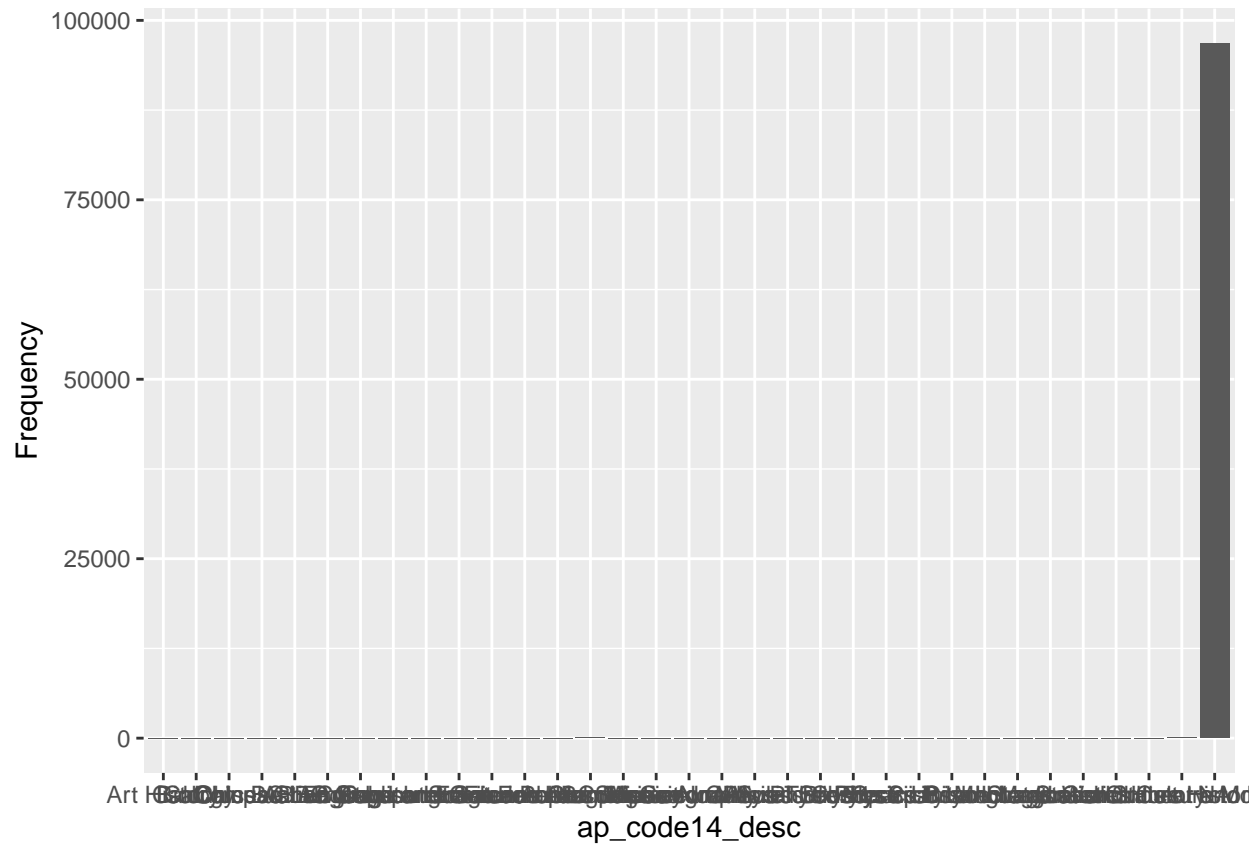
```
[1] Most missing: 2006 100%, Least missing: 2018 97%
```



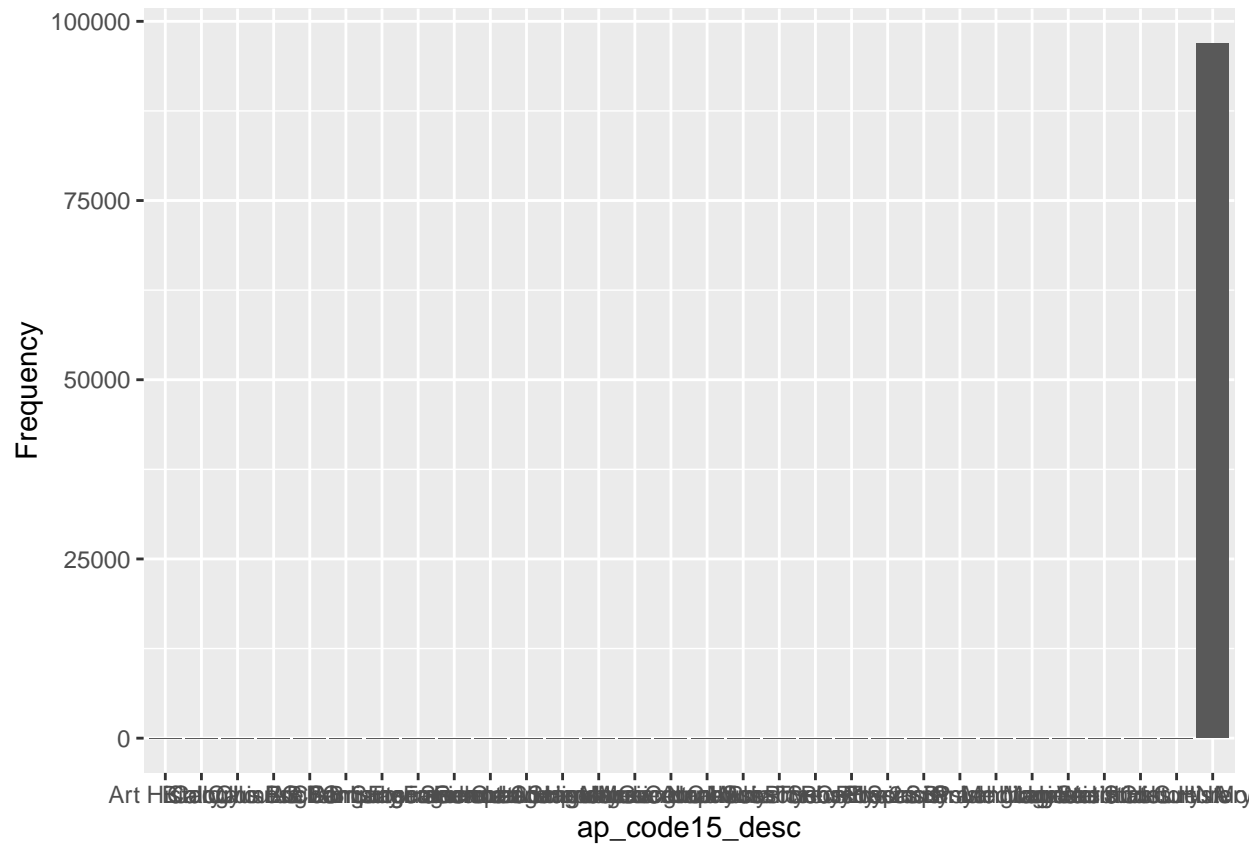
```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code13_desc, type: character, requested: no
[1] Values (31 unique): NA, World History: Modern, Calculus AB, Statistics, Chemistry, ...
[1] Missing: 99.4%
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code14_desc, type: character, requested: no
[1] Values (33 unique): NA, Physics 1, Biology, World History: Modern, Spanish Language and Culture, ..
[1] Missing: 99.7%
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code15_desc, type: character, requested: no
[1] Values (30 unique): NA, World History: Modern, Calculus BC, Biology, Music Aural Subscore, ...
[1] Missing: 99.9%
```



[1] should not be used as a predictor

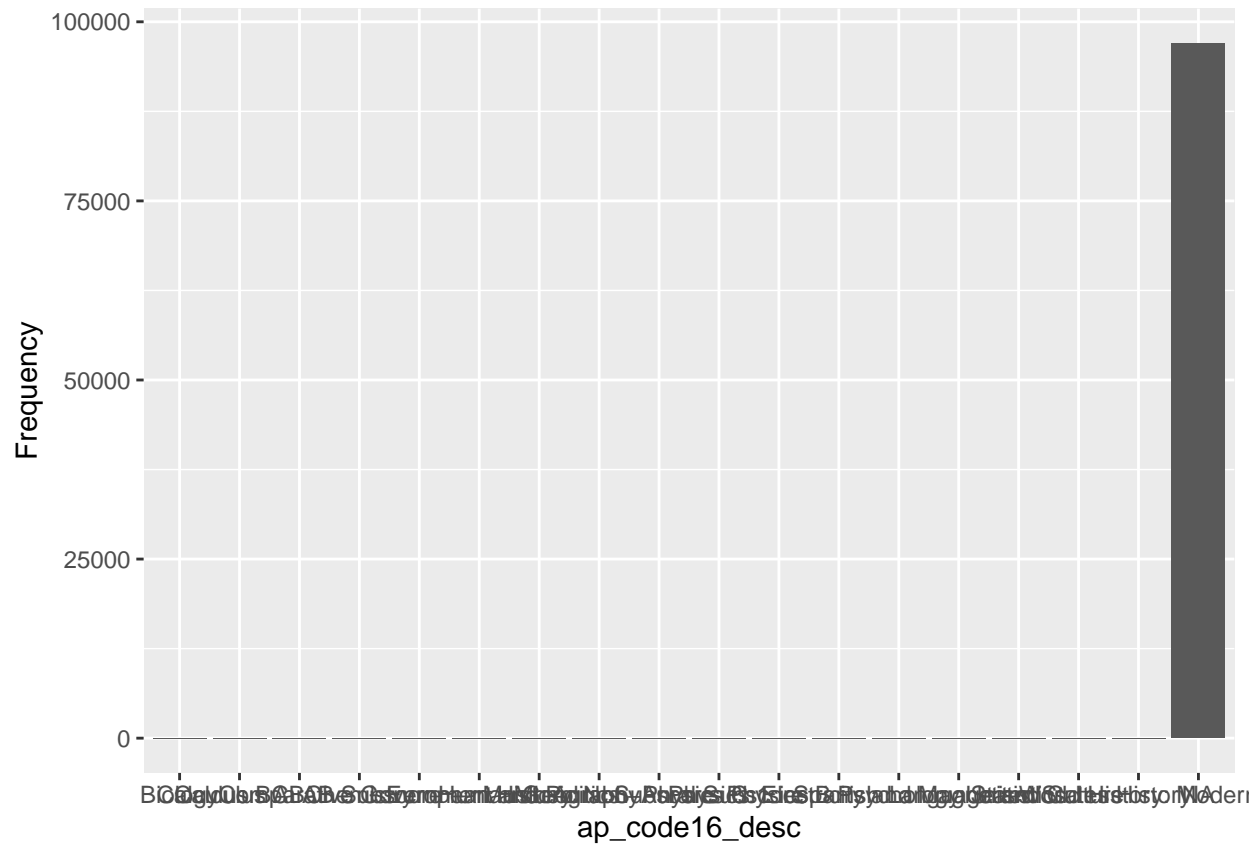
[1] is used in feature engineering

[1] -----

[1] Variable: ap_code16_desc, type: character, requested: no

[1] Values (18 unique): NA, Calculus BC: AB Subscore, European History, Music Non-Aural Subscore, World

[1] Missing: 99.9%



[1] should not be used as a predictor

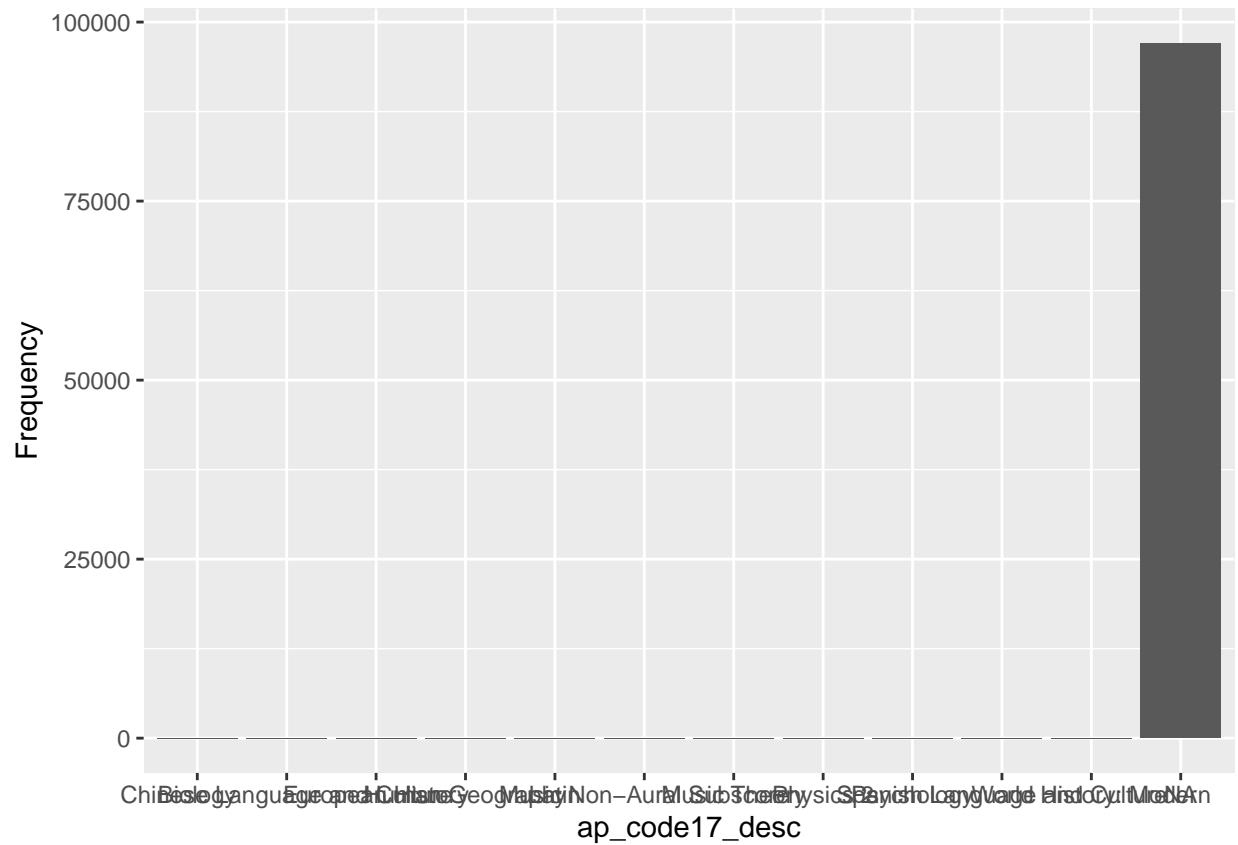
[1] is used in feature engineering

[1] _____

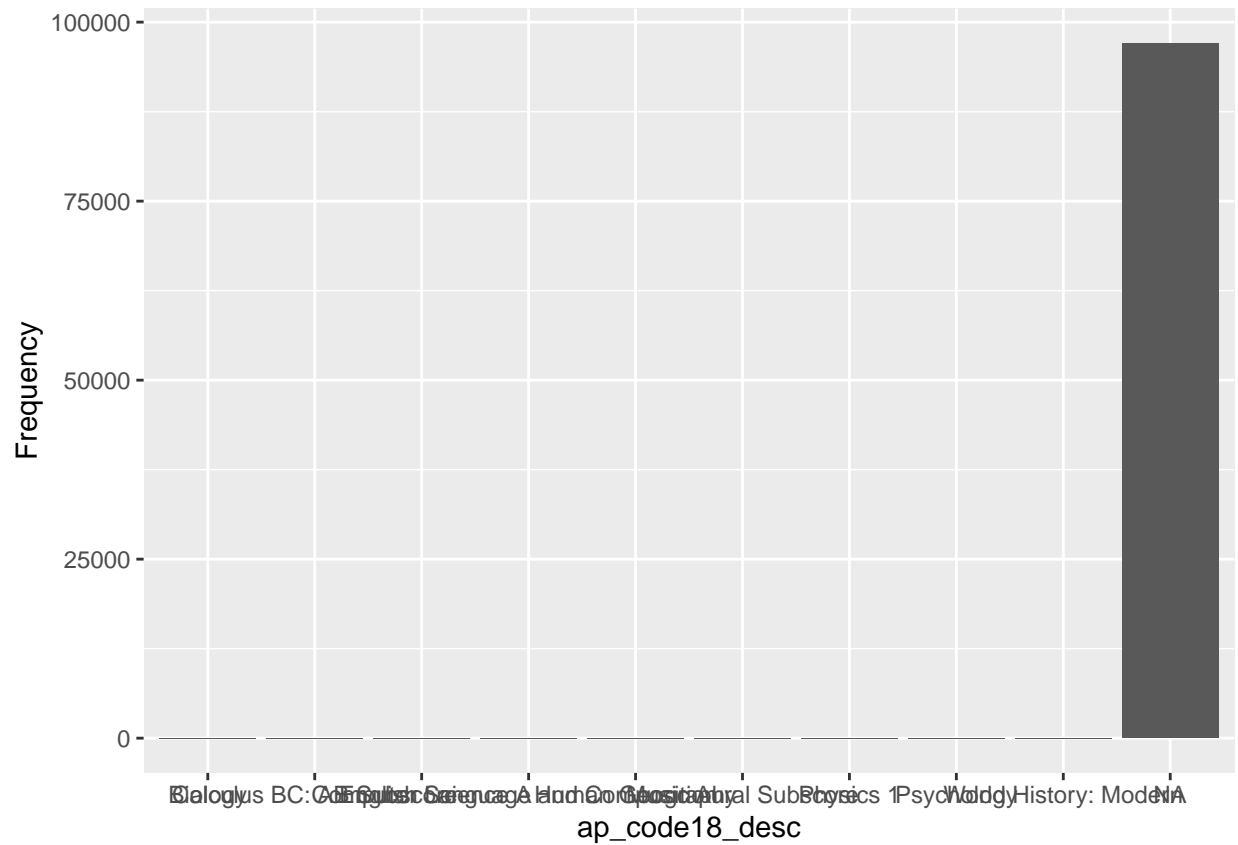
```
[1] Variable: ap_code17_desc, type: character, requested: no
```

[1] Values (12 unique): NA, World History: Modern, Chinese Language and Culture, Spanish Language and Culture, Chinese Language and Culture, Spanish Language and Culture, Chinese Language and Culture, Spanish Language and Culture, Chinese Language and Culture, Spanish Language and Culture, Chinese Language and Culture, Spanish Language and Culture

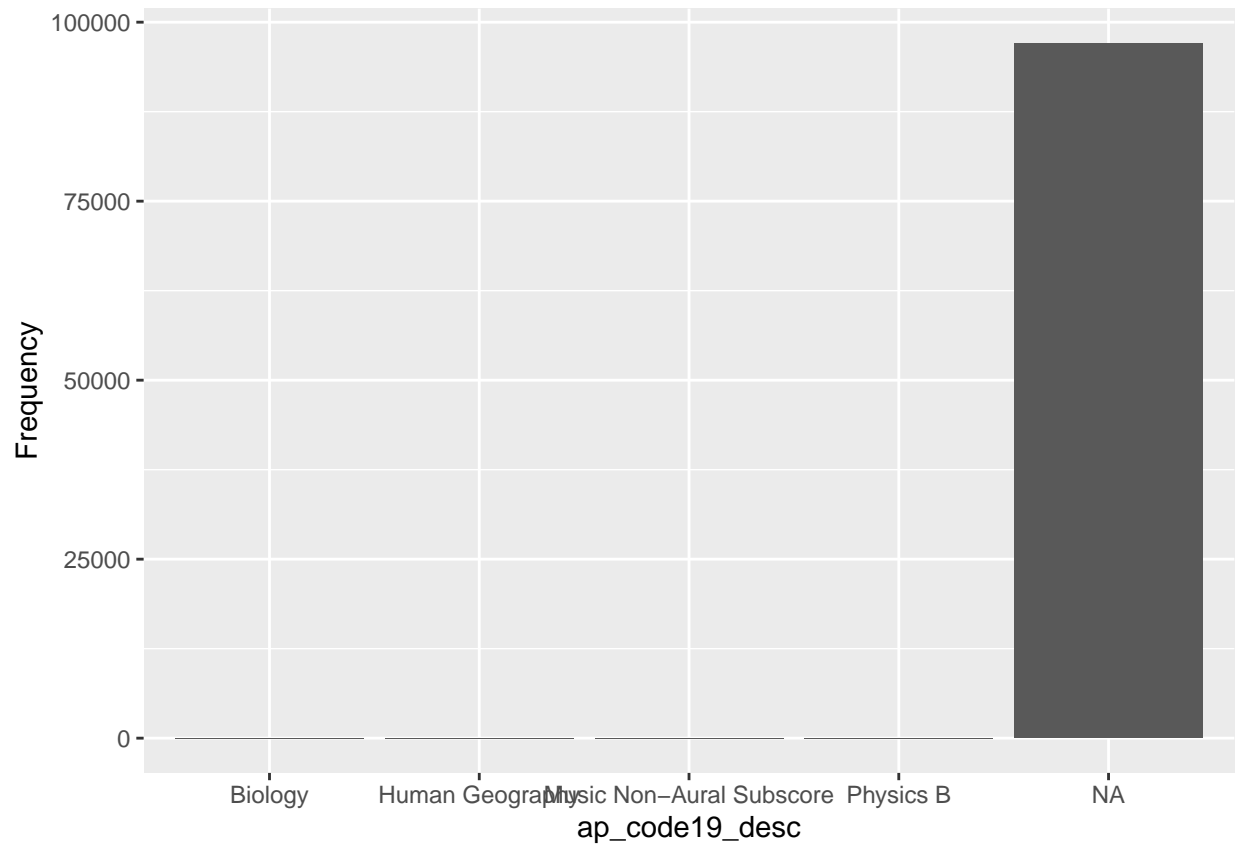
```
[1] Missing: 100%
```



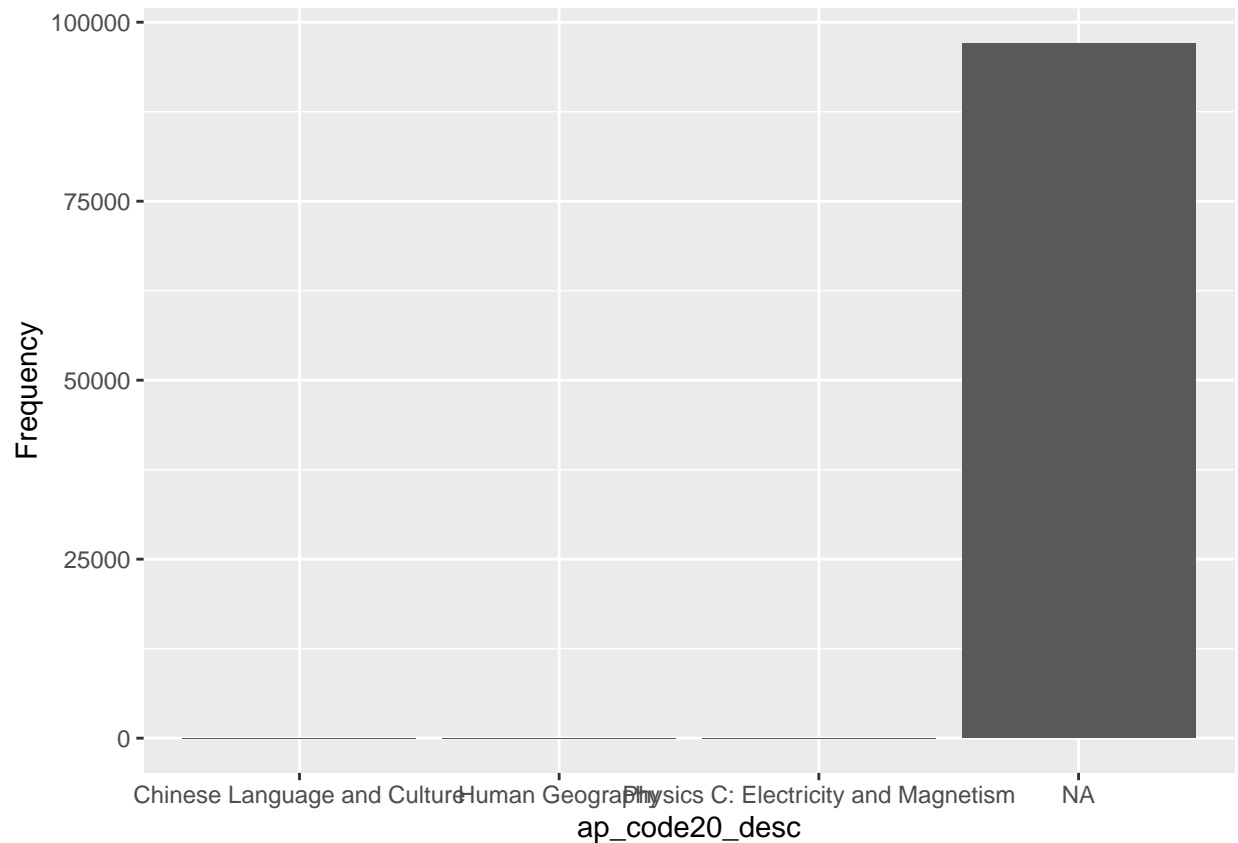
```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code18_desc, type: character, requested: no
[1] Values (10 unique): NA, Biology, World History: Modern, Human Geography, Psychology, ...
[1] Missing: 100%
```

```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code19_desc, type: character, requested: no
[1] Values (5 unique): NA, Human Geography, Biology, Music Non-Aural Subscore, Physics B
[1] Missing: 100%
```



```
[1] should not be used as a predictor
[1] is used in feature engineering
[1] -----
[1] Variable: ap_code20_desc, type: character, requested: no
[1] Values (4 unique): NA, Human Geography, Chinese Language and Culture, Physics C: Electricity and Mag
[1] Missing: 100%
```



```
[1] should not be used as a predictor
```

```
[1] is used in feature engineering
```

```
[1] -----
```

```
[1] Variable: uncapped_gpa, type: numeric, requested: no
```

```
[1] Values (1 unique): NA
```

```
[1] Missing: 100%
```

```
Warning in min(x): kein nicht-fehlendes Argument für min; gebe Inf zurück
```

```
Warning in max(x): kein nicht-fehlendes Argument für max; gebe -Inf zurück
```

```
Warning in min(diff(sort(x))): kein nicht-fehlendes Argument für min; gebe Inf zurück
```

```
Warning: Removed 97049 rows containing non-finite values ('stat_count()').
```

Frequency

uncapped_gpa

```
[1] -----  
[1] Variable: unweighted_gpa, type: numeric, requested: no  
[1] Values (1 unique): NA  
[1] Missing: 100%
```

Warning in min(x): kein nicht-fehlendes Argument für min; gebe Inf zurück

Warning in max(x): kein nicht-fehlendes Argument für max; gebe -Inf zurück

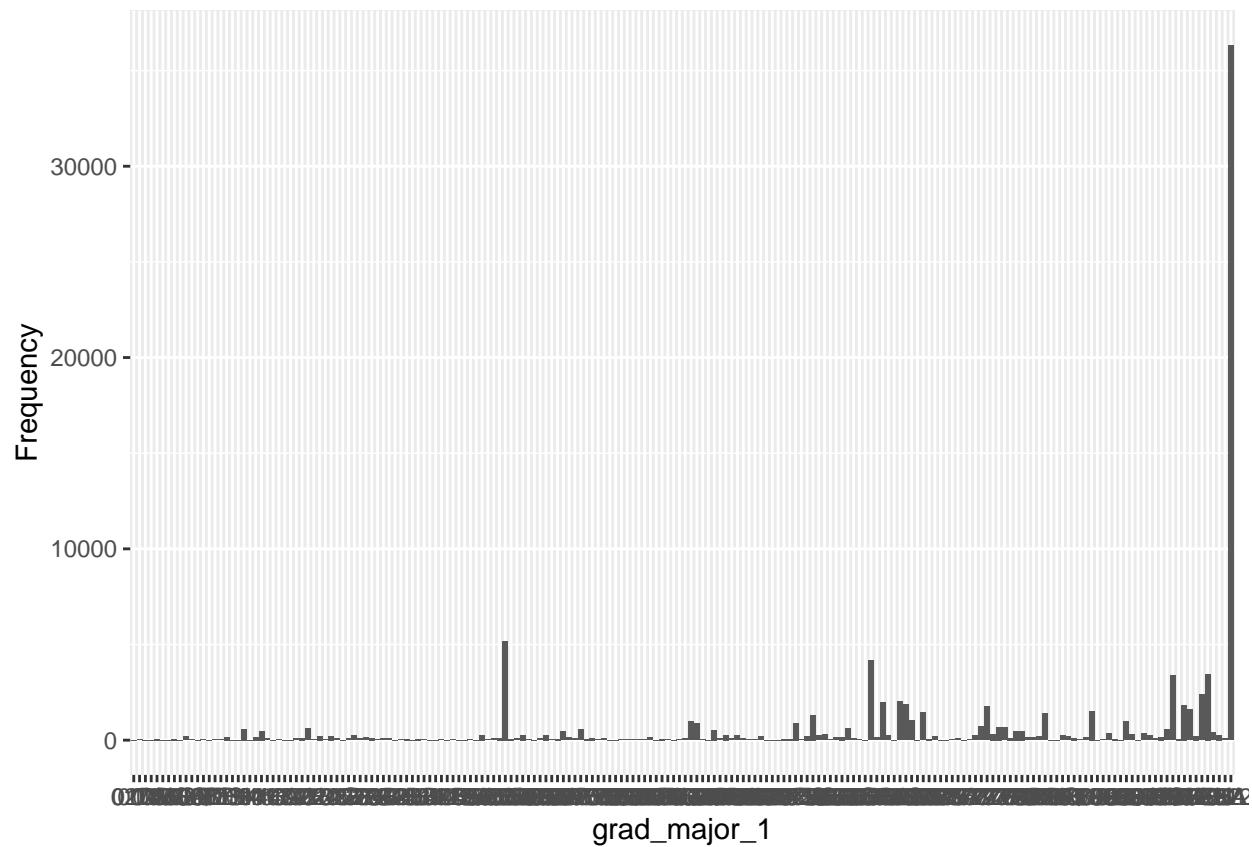
Warning in min(diff(sort(x))): kein nicht-fehlendes Argument für min; gebe Inf zurück

Warning: Removed 97049 rows containing non-finite values ('stat_count()').

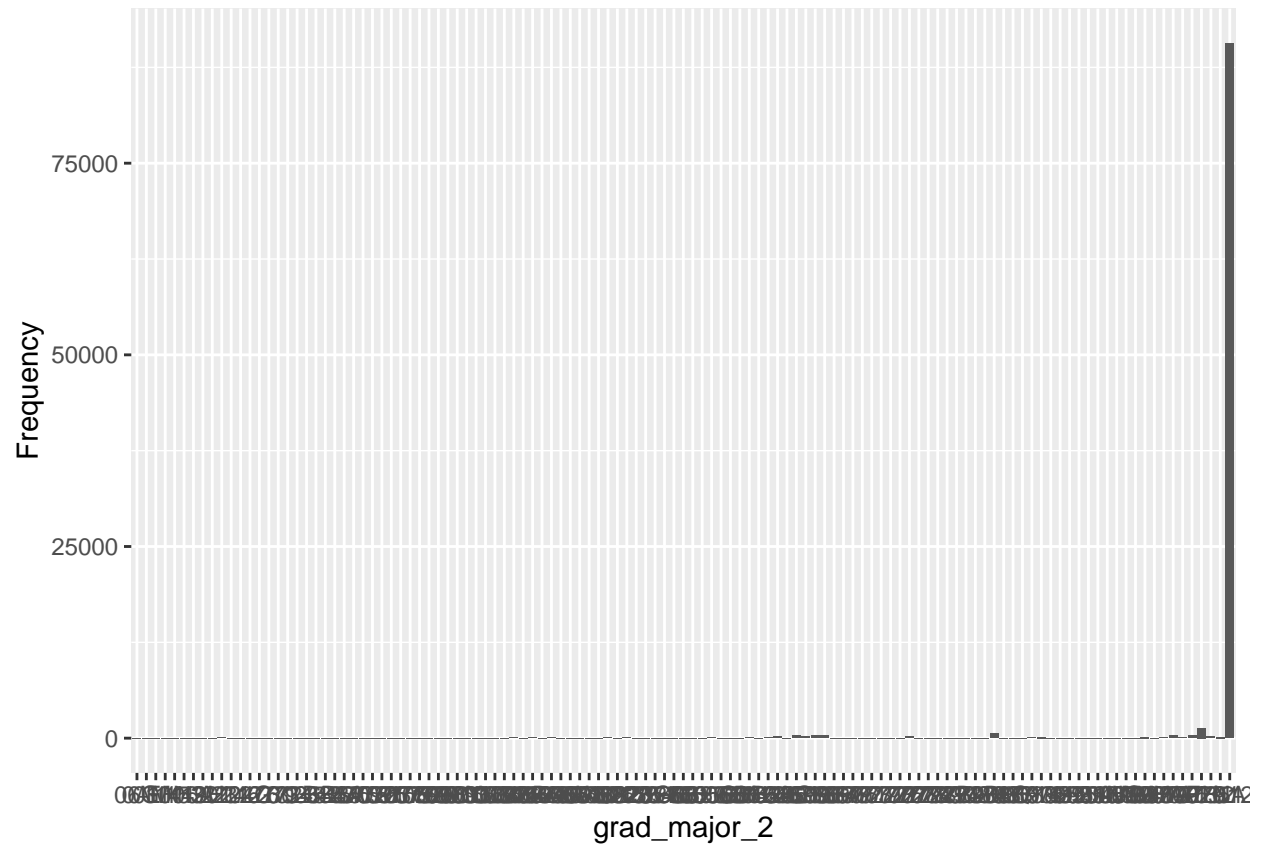
Frequency

unweighted_gpa

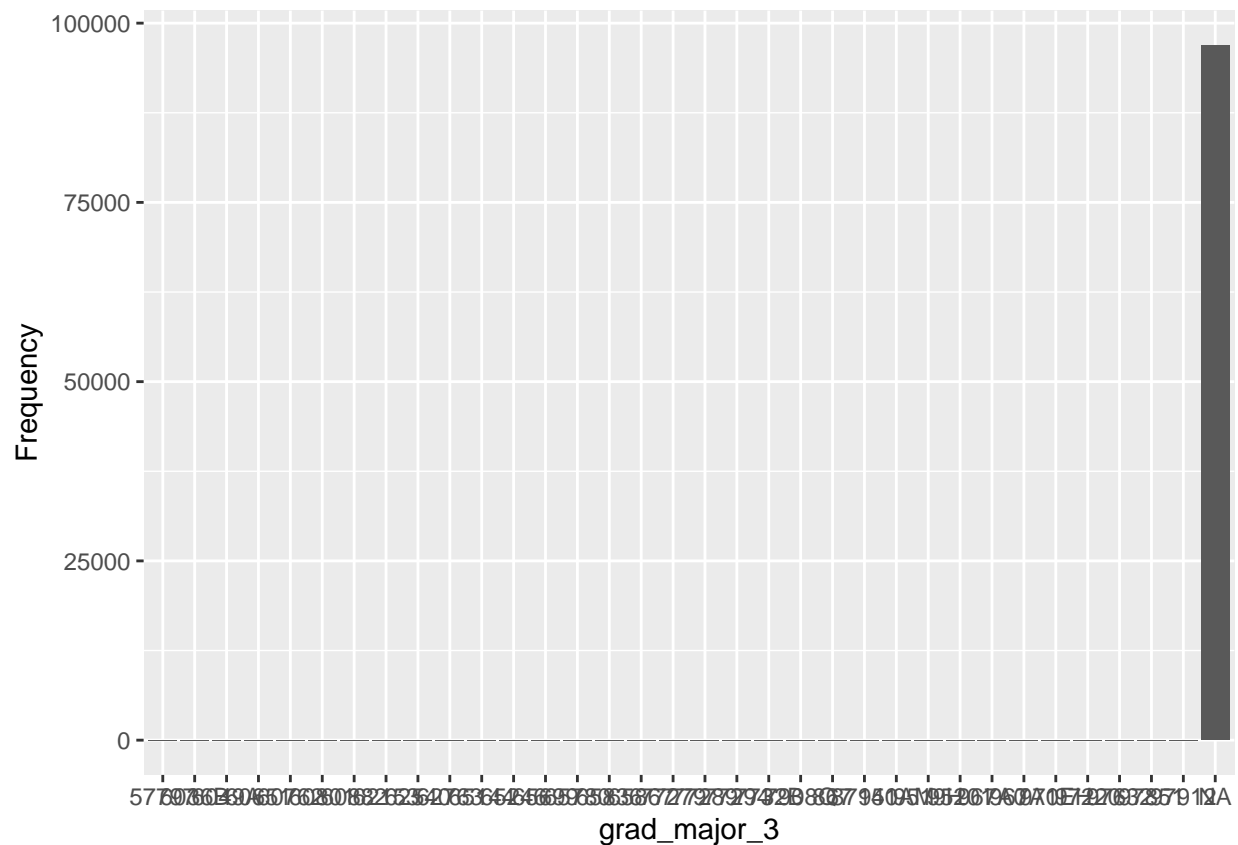
```
[1] -----
[1] Variable: grad_major_1, type: character, requested: no
[1] Values (190 unique): NA, 55120, 62153, 65780, 65144, ...
[1] Missing: 37.4%
  Group.1 grad_major_1
1      2006      0.9978039
2      2007      0.9974945
3      2008      0.9964621
4      2009      0.9956369
5      2010      0.1387730
6      2011      0.1485836
7      2012      0.1613042
8      2013      0.1496590
9      2014      0.1549479
10     2015      0.1602370
11     2016      0.1355367
12     2017      0.1299784
13     2018      0.1823096
```



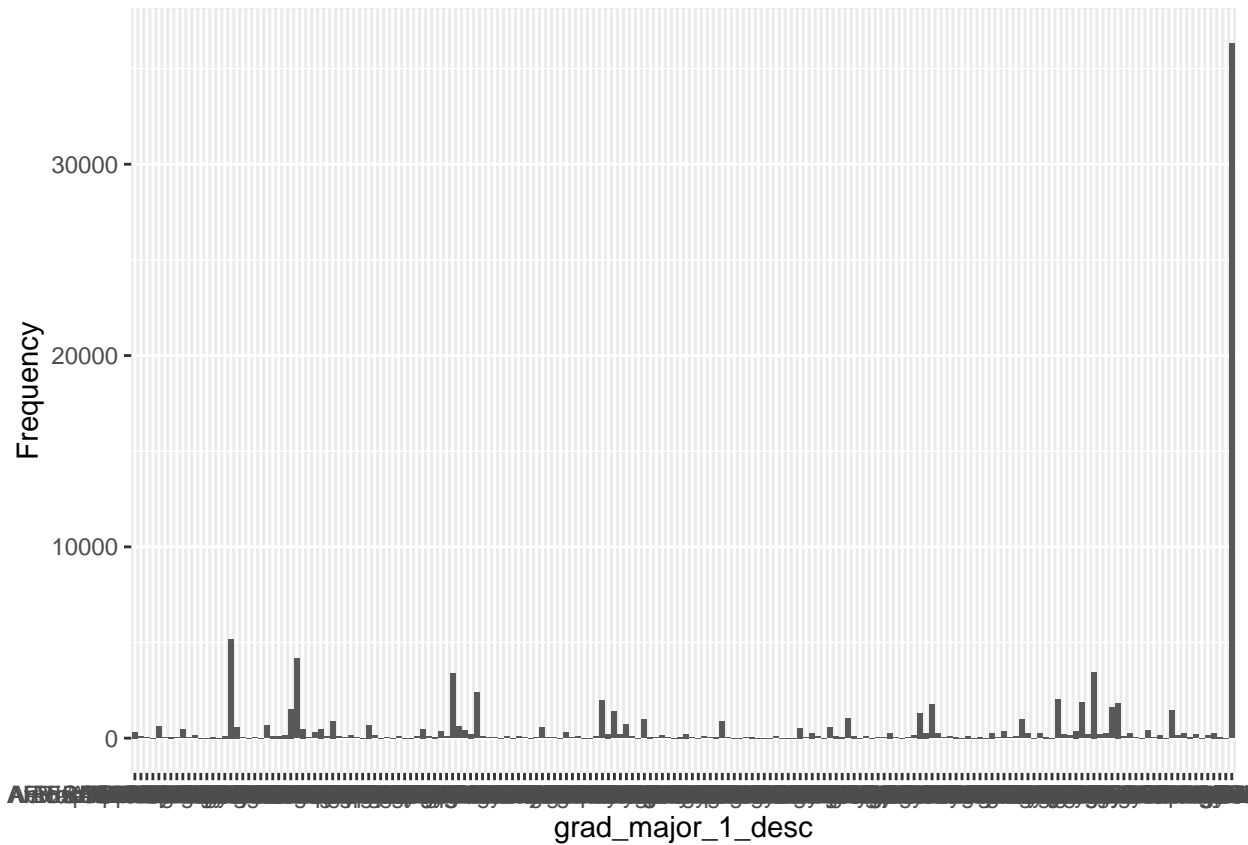
```
[1] -----
[1] Variable: grad_major_2, type: character, requested: no
[1] Values (117 unique): NA, 65836, 65246, 97220, 967A1, ...
[1] Missing: 93.4%
[1] Most missing: 2006 100%, Least missing: 2011 88.6%
```



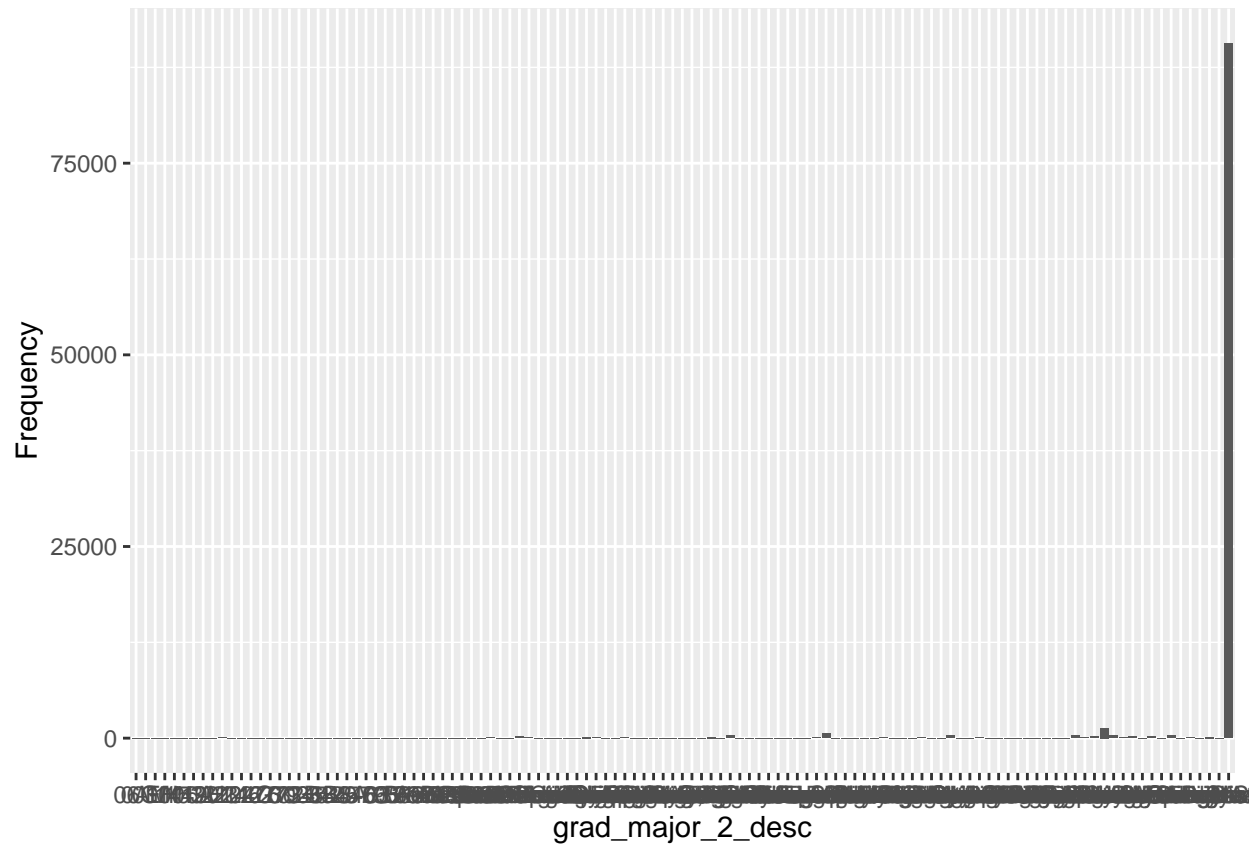
```
[1] -----
[1] Variable: grad_major_3, type: character, requested: no
[1] Values (34 unique): NA, 97220, 970EH, 97632, 65246, ...
[1] Missing: 99.8%
```



```
[1] -----
[1] Variable: grad_major_1_desc, type: character, requested: no
[1] Values (184 unique): NA, Biological Sciences, Chemistry, Psychology, Business Economics, ...
[1] Missing: 37.4%
  Group.1 grad_major_1_desc
1    2006      0.9978039
2    2007      0.9974945
3    2008      0.9964621
4    2009      0.9956369
5    2010      0.1387730
6    2011      0.1485836
7    2012      0.1613042
8    2013      0.1496590
9    2014      0.1549479
10   2015      0.1602370
11   2016      0.1355367
12   2017      0.1299784
13   2018      0.1823096
```

```
[1] should not be used as a predictor
[1] -----
[1] Variable: grad_major_2_desc, type: character, requested: no
[1] Values (115 unique): NA, International Studies, Economics, Criminology, Law And Society, Public Hea
[1] Missing: 93.4%
[1] Most missing: 2006 100%, Least missing: 2011 88.6%
```



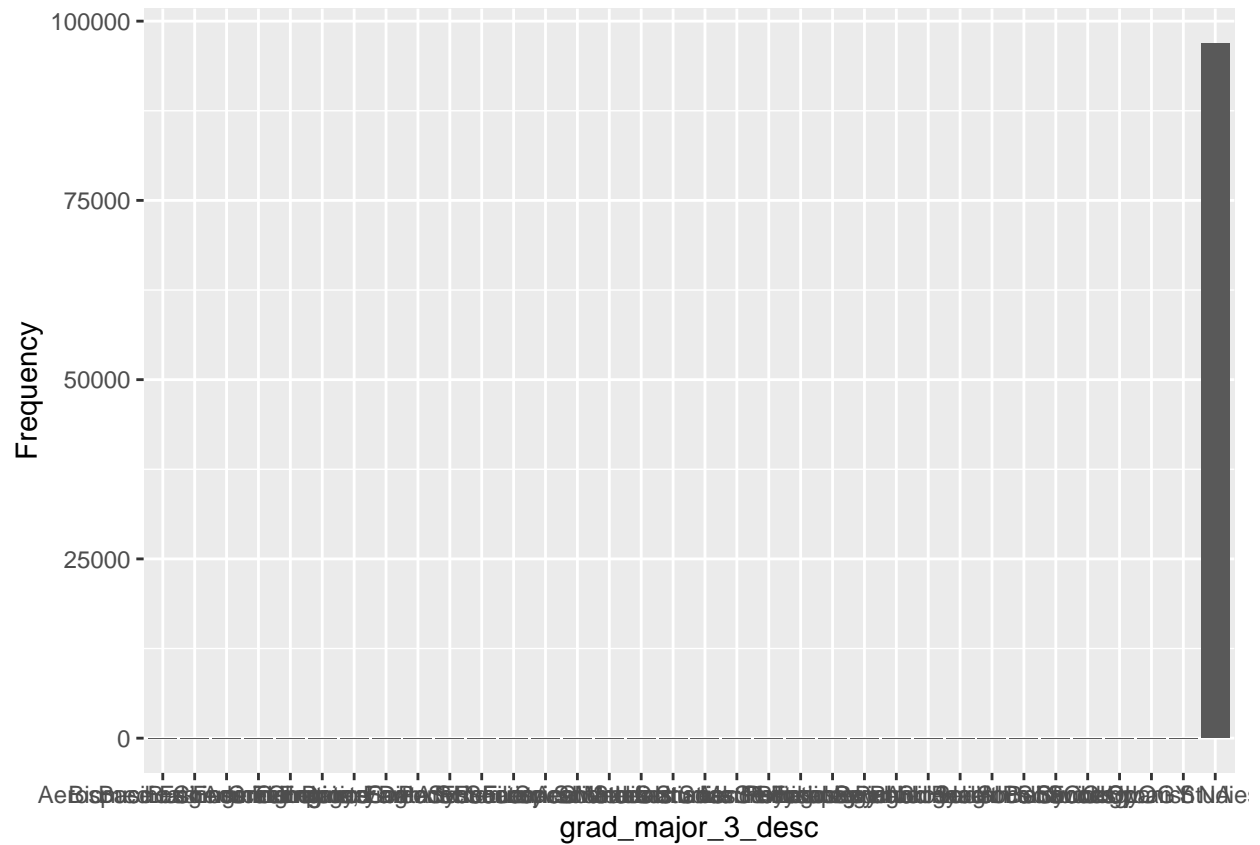
[1] should not be used as a predictor

[1] _____

```
[1] Variable: grad_major_3_desc, type: character, requested: no
```

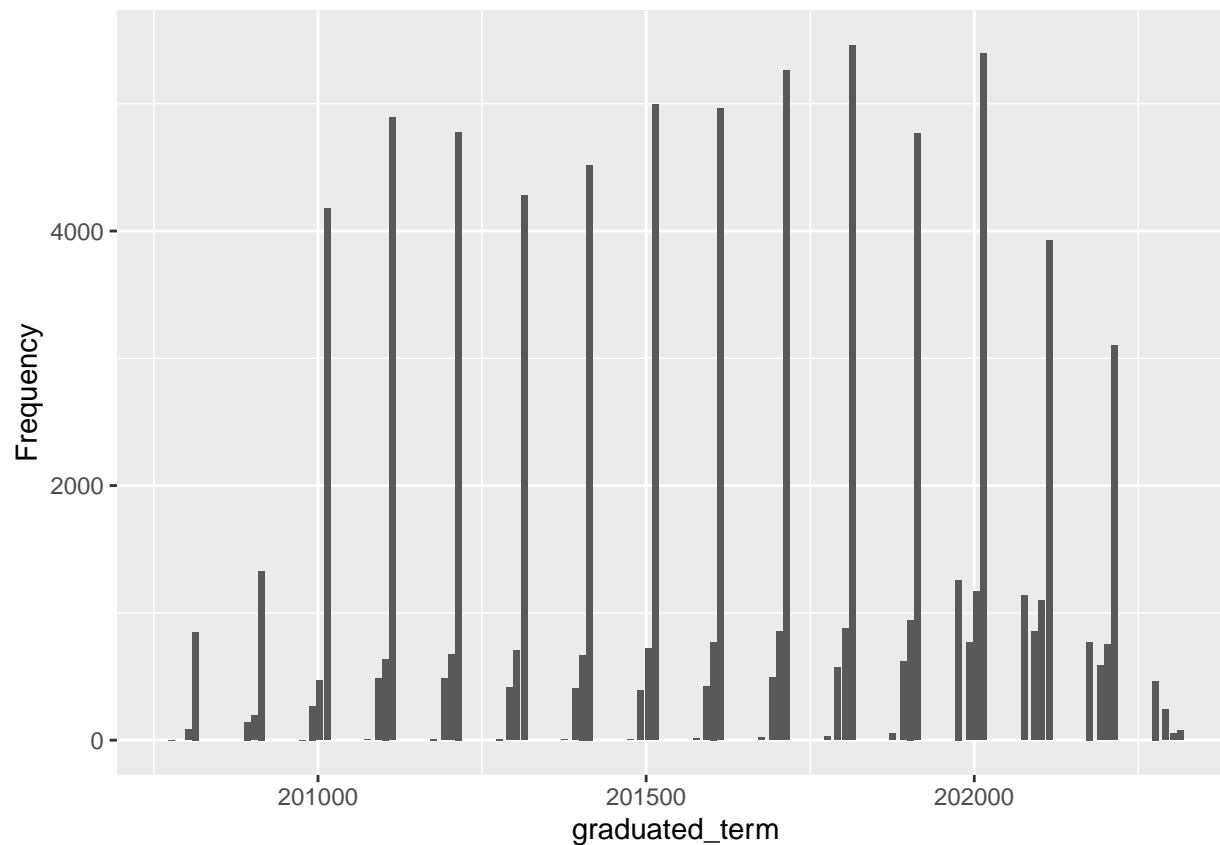
[1] Values (34 unique): NA, Criminology, Law And Society, Psychological Science, Psychology And Social

```
[1] Missing: 99.8%
```



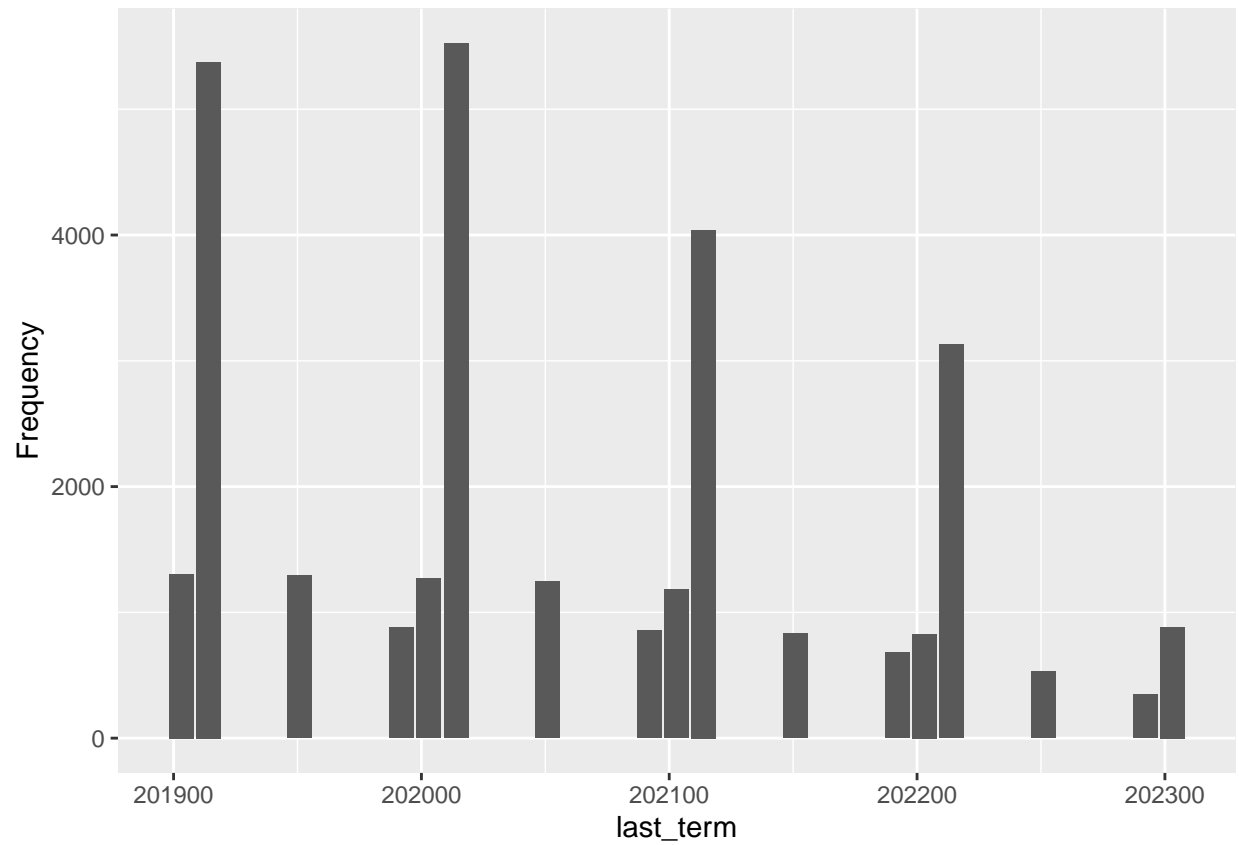
```
[1] should not be used as a predictor
[1] -----
[1] Variable: graduated_term, type: numeric, requested: no
[1] Values (63 unique): NA, 201914, 202076, 202014, 202114, ...
[1] Missing: 13.1%
[1] Most missing: 2018 18.2%, Least missing: 2006 5.6%
```

Warning: Removed 12701 rows containing non-finite values ('stat_count()').

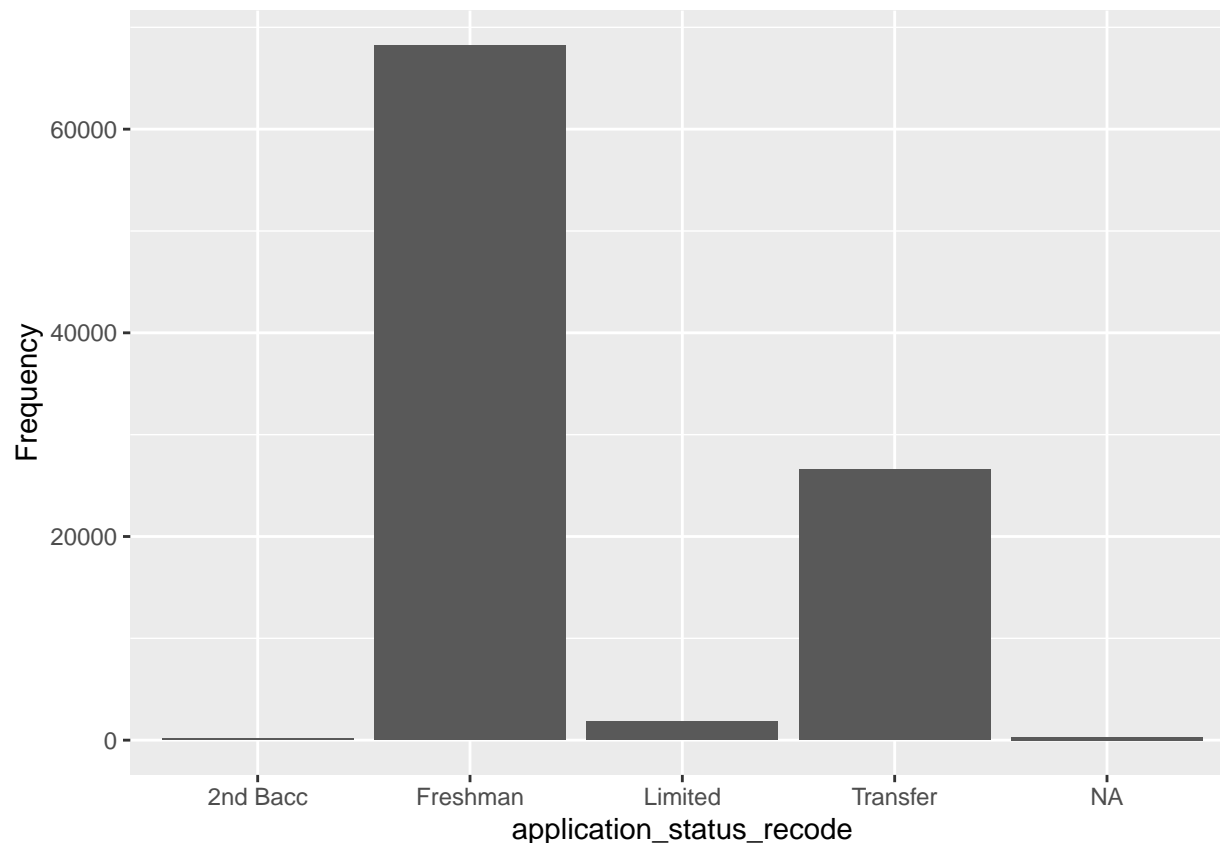


```
[1] -----
[1] Variable: last_term, type: numeric, requested: no
[1] Values (18 unique): 202092, 201914, 201903, 202014, 201951, ...
[1] Missing: 68.9%
Group.1 last_term
1      2006 0.99701961
2      2007 0.99749447
3      2008 0.99630826
4      2009 0.99496560
5      2010 0.99613661
6      2011 0.99079320
7      2012 0.98850905
8      2013 0.97520986
9      2014 0.88776042
10     2015 0.41805345
11     2016 0.27290343
12     2017 0.08922507
13     2018 0.00310559
```

Warning: Removed 66851 rows containing non-finite values ('stat_count()').



```
[1] -----
[1] Variable: application_status_recode, type: character, requested: no
[1] Values (5 unique): Transfer, NA, Freshman, Limited, 2nd Bacc
[1] Missing: 0.3%
```



```
[1] -----
```

These are the predictors that remain when filtering for less than 50% missing, our blacklist and variables that need to be somehow transformed first:

```
[1] "group_a"           "group_b"
[3] "gender"            "female"
[5] "int_student"       "ethnicity"
[7] "asian"             "hispanic"
[9] "black"             "white"
[11] "urm"               "citizenship_app"
[13] "first_generation"  "low_income"
[15] "low_income_desc"   "father_edu_level_code"
[17] "father_edu_level_desc" "mother_edu_level_code"
[19] "mother_edu_level_desc" "ell"
[21] "foster_care"       "single_parent"
[23] "geo_category"      "application_status"
[25] "cal_res_at_app"    "current_ind"
[27] "parent_educ"       "pascd"
[29] "application_status_recode"
```

```
write.csv2(bg_vars, 'bg_vars.csv', )
```