Predicting Earthquake Building Damage

Haley Egan

Nepal

2015

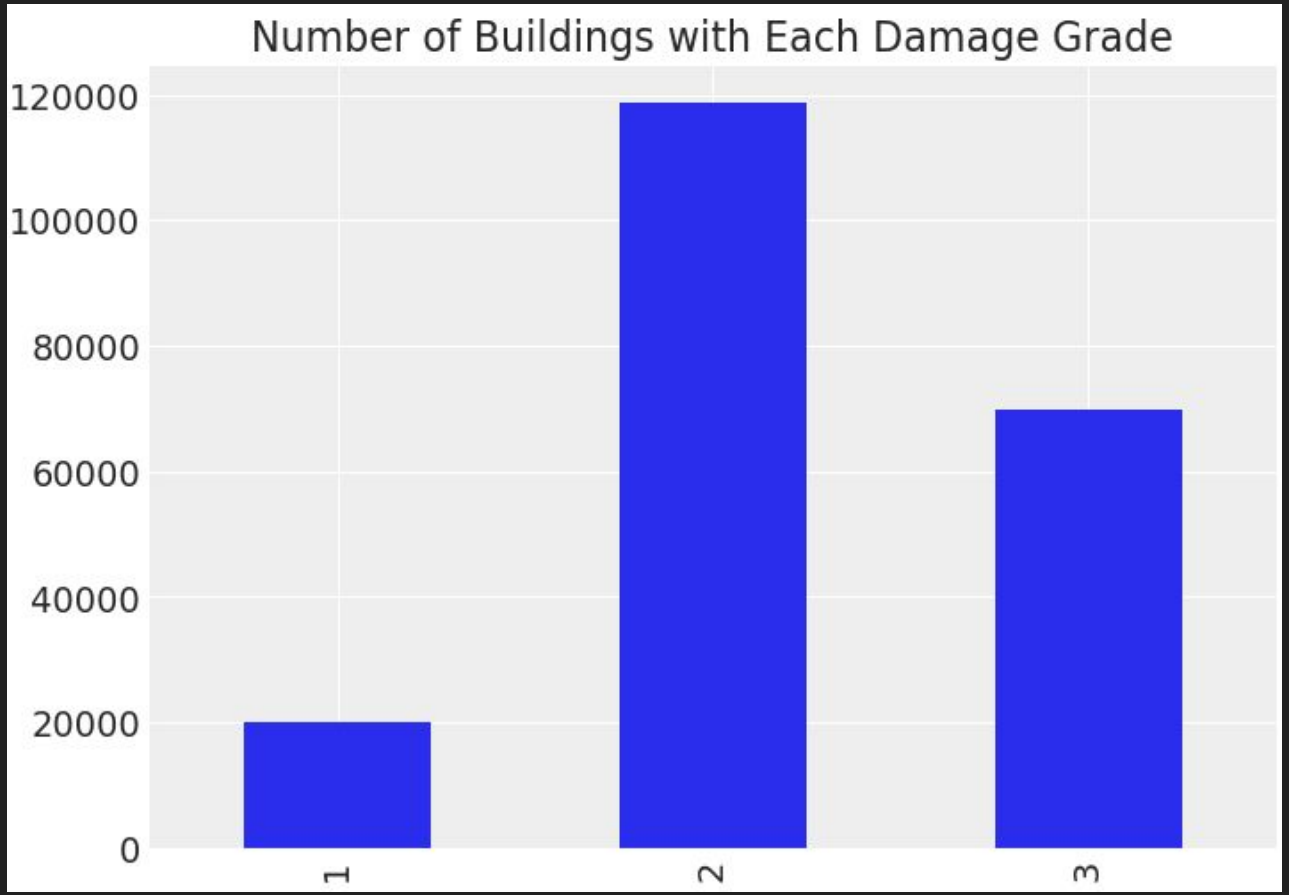
Data

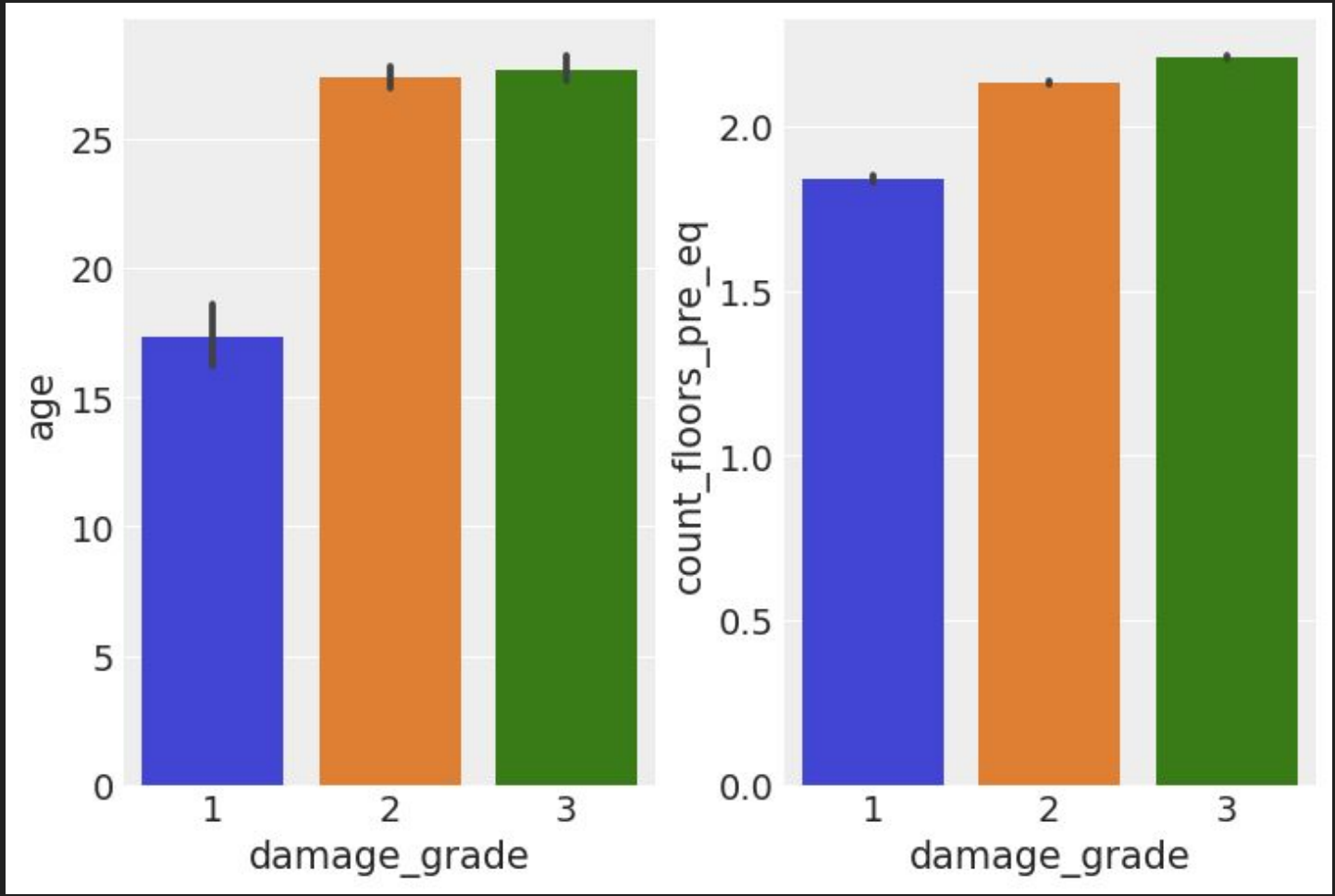
- 2015 Nepal Earthquake Open Data Portal
- National Planning Commission, Kathmandu Living Labs, and the Central Bureau of Statistics
- 208,480 buildings (rows)
- 38 predictors, mixed types (columns)
- Damage Grade of each building
  - 1 = low damage
  - 2 = medium damage
  - 3 = complete destruction

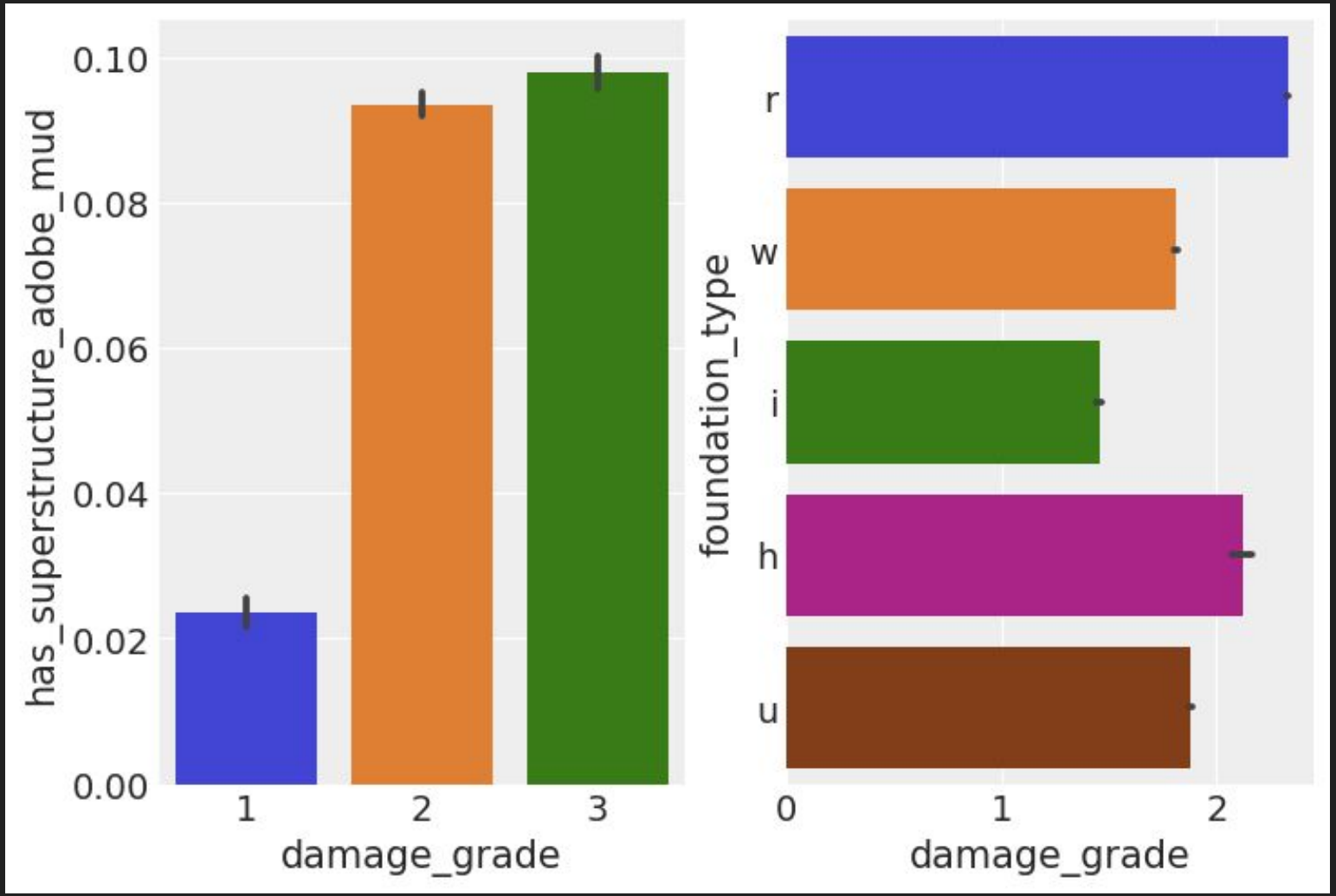
Priors

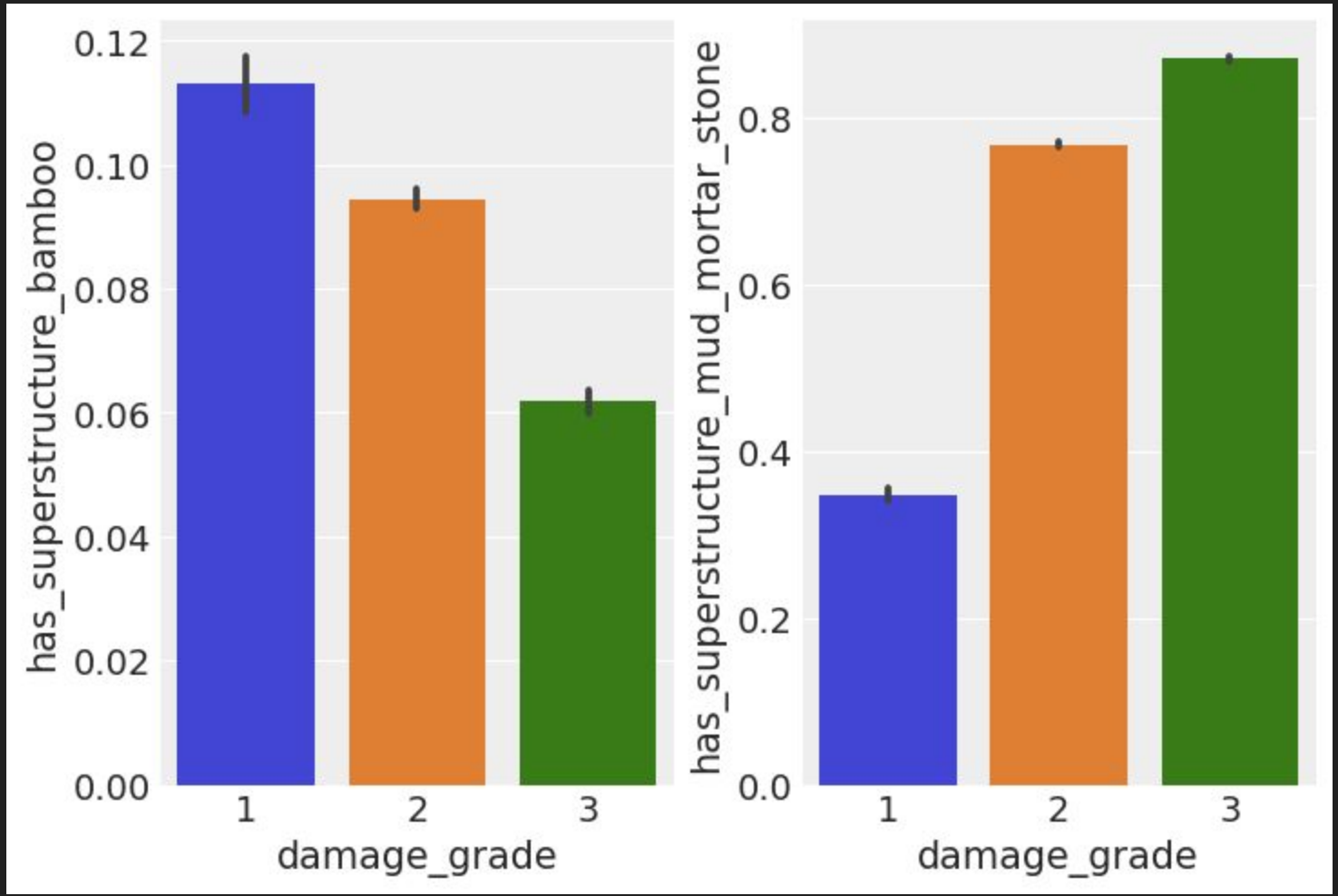
- Priors from knowledge/observations
  - Most buildings had some damage, some had severe damage
  - Damage grade 1 = 5%, grade 2 = 70%, grade 3 = 25%
- Priors from observed proportion
  - Buildings per grade/total buildings
  - Damage grade 1 = 9.7%, grade 2 = 57%, grade 3 = 33.4%

Exploratory Data Analysis




age
25
20
15
10
5
0
1
2
3
damage_grade
count_floors_pre_eq
2.0
1.5
1.0
0.5
0.0
1
2
3
damage_grade

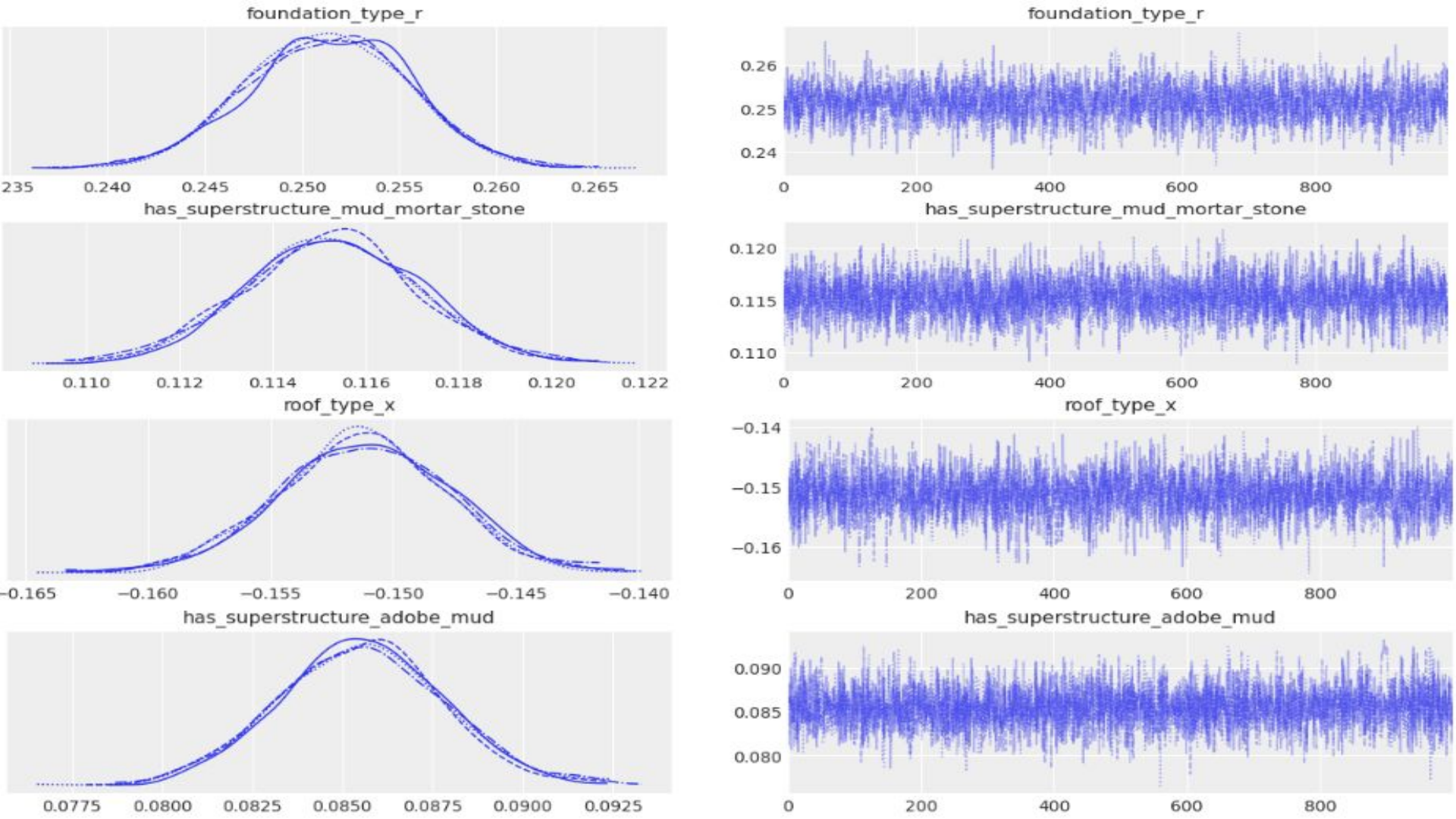

has_superstructure_adobe_mud
0.10
0.08
0.06
0.04
0.02
0.00
1
2
3
damage_grade
foundation_type
r
w
i
h
u
0
1
2
damage_grade

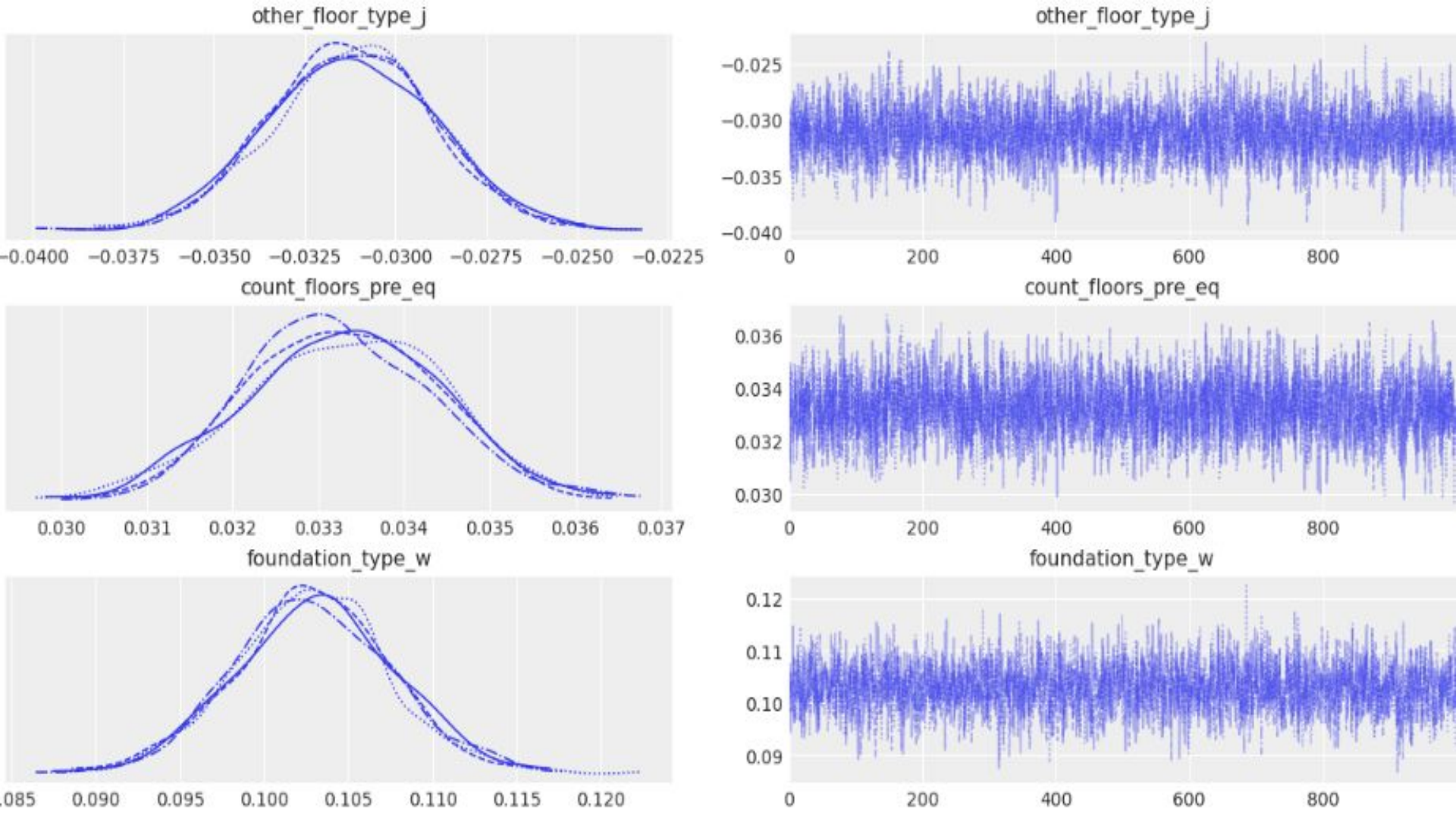

has_superstructure_bamboo
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1
2
3
damage_grade
has_superstructure_mud_mortar_stone
0.8
0.6
0.4
0.2
0.0
1
2
3
damage_grade

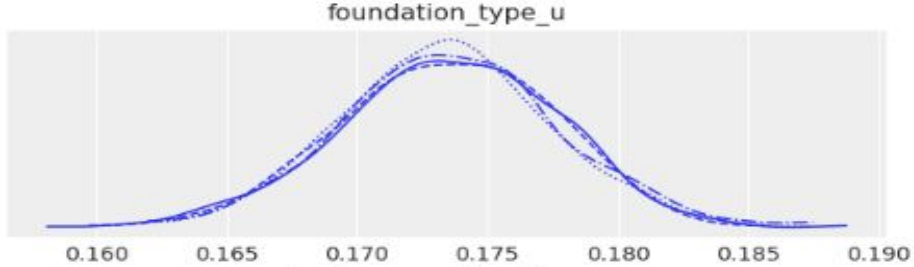
Models

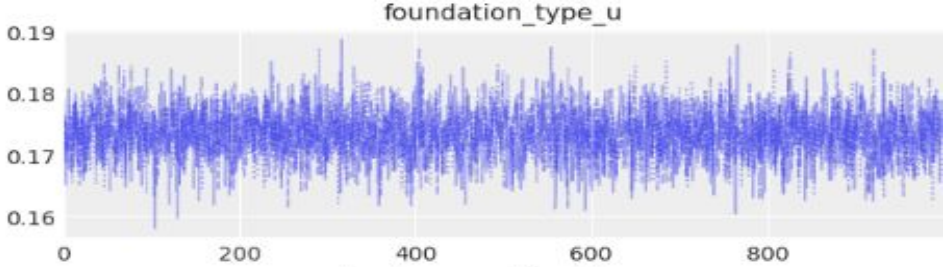
- Linear Discriminant Analysis (LDA)
- Quadratic Discriminant Analysis (QDA)
- Logistic Regression
- MCMC, trace plots, density plots

Accuracy	Training Data	Testing Data
LDA	43.9%	11.1%
QDA	25.1%	6.5%
Logistic Regression	90.9%	90.8%
BMA	90.9%	90.8%

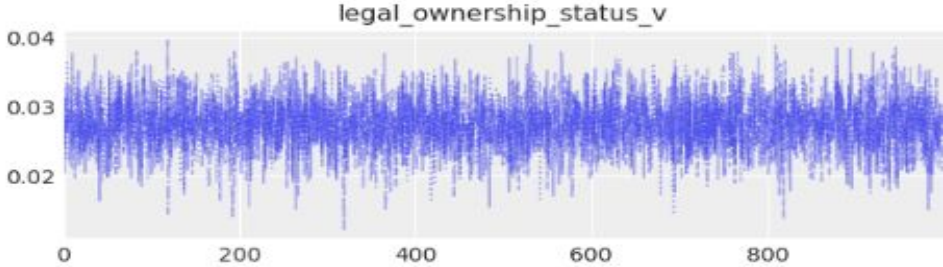

foundation_type_r
235
0.240
0.245
0.250
0.255
0.260
0.265
foundation_type_r
0.26
0.25
0.24
0
200
400
600
800
has_superstructure_mud_mortar_stone
0.110
0.112
0.114
0.116
0.118
0.120
0.122
has_superstructure_mud_mortar_stone
0.120
0.115
0.110
0
200
400
600
800
roof_type_x
−0.165
−0.160
−0.155
−0.150
−0.145
−0.140
roof_type_x
−0.14
−0.15
−0.16
0
200
400
600
800
has_superstructure_adobe_mud
0.0775
0.0800
0.0825
0.0850
0.0875
0.0900
0.0925
has_superstructure_adobe_mud
0.090
0.085
0.080
0
200
400
600
800

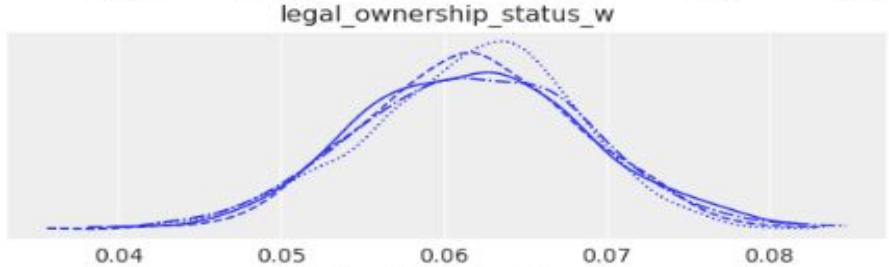

other_floor_type_j
−0.0400 −0.0375 −0.0350 −0.0325 −0.0300 −0.0275 −0.0250 −0.0225
other_floor_type_j
−0.025
−0.030
−0.035
−0.040
0 200 400 600 800
count_floors_pre_eq
0.030 0.031 0.032 0.033 0.034 0.035 0.036 0.037
count_floors_pre_eq
0.036
0.034
0.032
0.030
0 200 400 600 800
foundation_type_w
.085 0.090 0.095 0.100 0.105 0.110 0.115 0.120
foundation_type_w
0.12
0.11
0.10
0.09
0 200 400 600 800

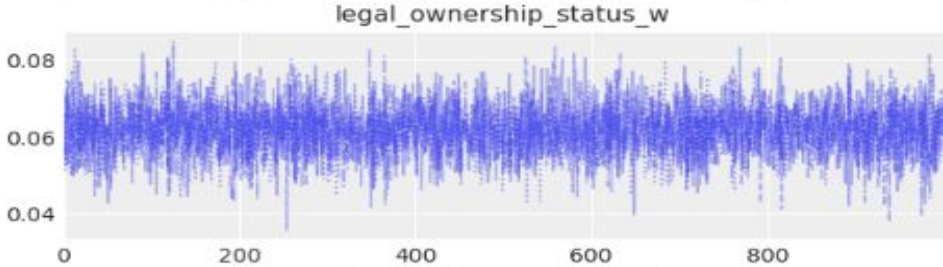

foundation_type_u
0.160
0.165
0.170
0.175
0.180
0.185
0.190

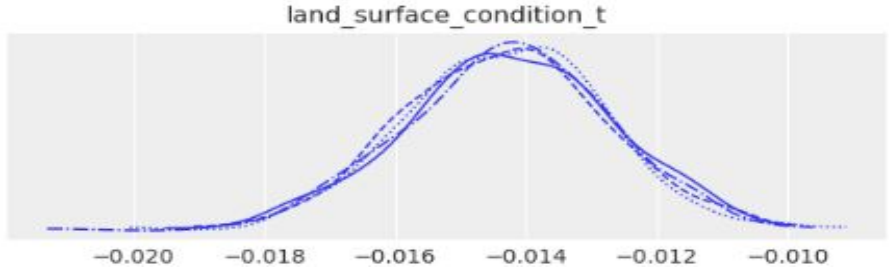

foundation_type_u
0.19
0.18
0.17
0.16
0
200
400
600
800

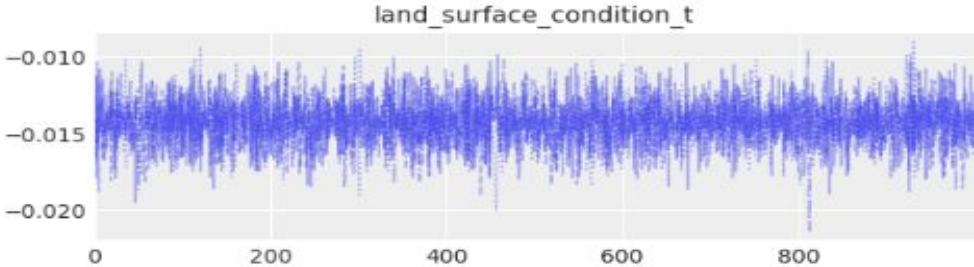

legal_ownership_status_v
0.015
0.020
0.025
0.030
0.035
0.040


legal_ownership_status_v
0.04
0.03
0.02
0
200
400
600
800


legal_ownership_status_w
0.04
0.05
0.06
0.07
0.08


legal_ownership_status_w
0.08
0.06
0.04
0
200
400
600
800


land_surface_condition_t
−0.020
−0.018
−0.016
−0.014
−0.012
−0.010


land_surface_condition_t
−0.010
−0.015
−0.020
0
200
400
600
800

Conclusions

- Logistic regression performs best
- QDA and LDA should not be used for predictions
- Further model explorations recommended
- Apply to other similar case studies