Multi-Level Model

Jack Welsh

May 1, 2019

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
## v ggplot2 3.1.0       v purrr   0.3.2
## v tibble  2.1.1       v dplyr   0.8.0.1
## v tidyr   0.8.3       v stringr 1.4.0
## v readr   1.3.1       v forcats 0.4.0
## -- Conflicts ------------------------------------------------------------------------------------ tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:base':
##
##     date
```

```
library(lme4)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following object is masked from 'package:tidyr':
##
##     expand
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(knitr)
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
adoptions <- read_csv("adoptions.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

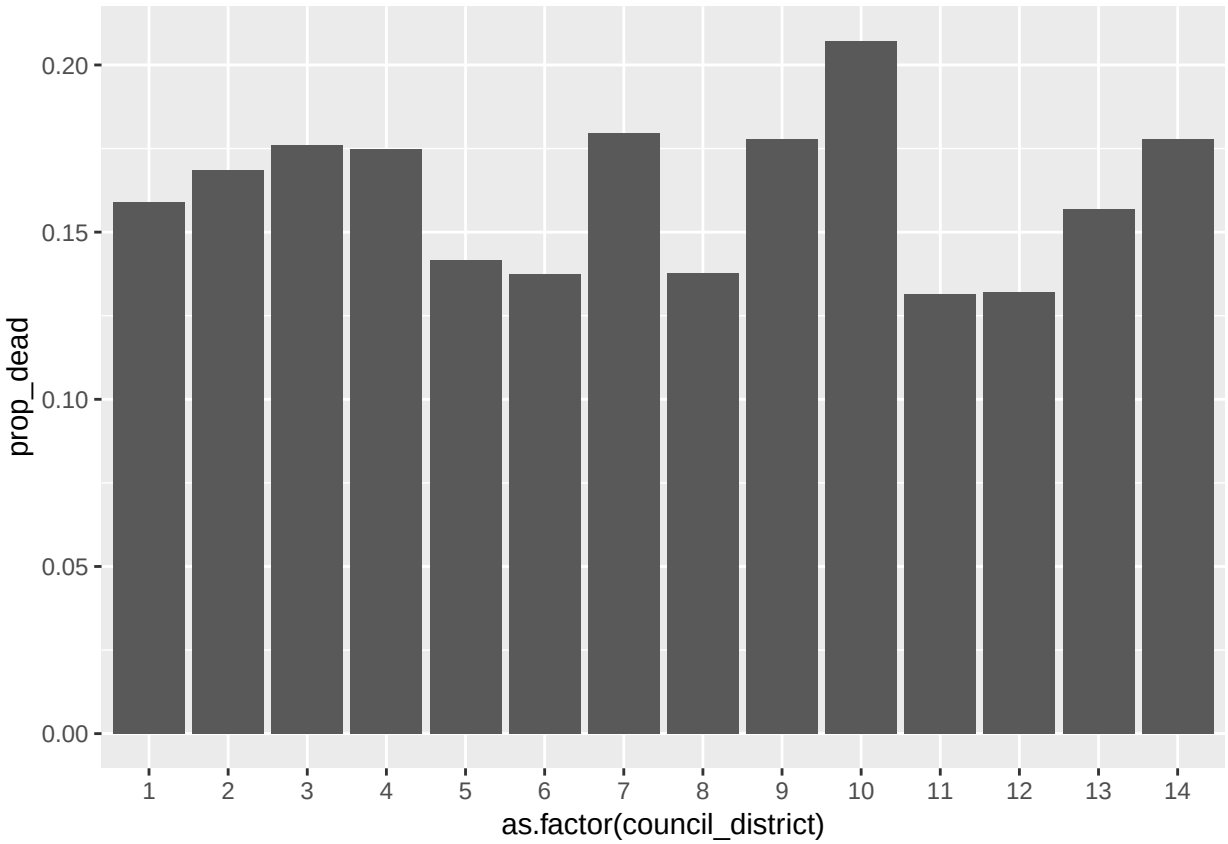
```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   animal_breed = col_character(),
##   animal_origin = col_character(),
##   animal_type = col_character(),
##   chip_status = col_character(),
##   intake_condition = col_character(),
##   intake_date = col_date(format = ""),
##   intake_subtype = col_character(),
##   intake_type = col_character(),
##   outcome_condition = col_character(),
##   outcome_date = col_date(format = ""),
##   outcome_type = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

Mulltilevel Logistic Regression

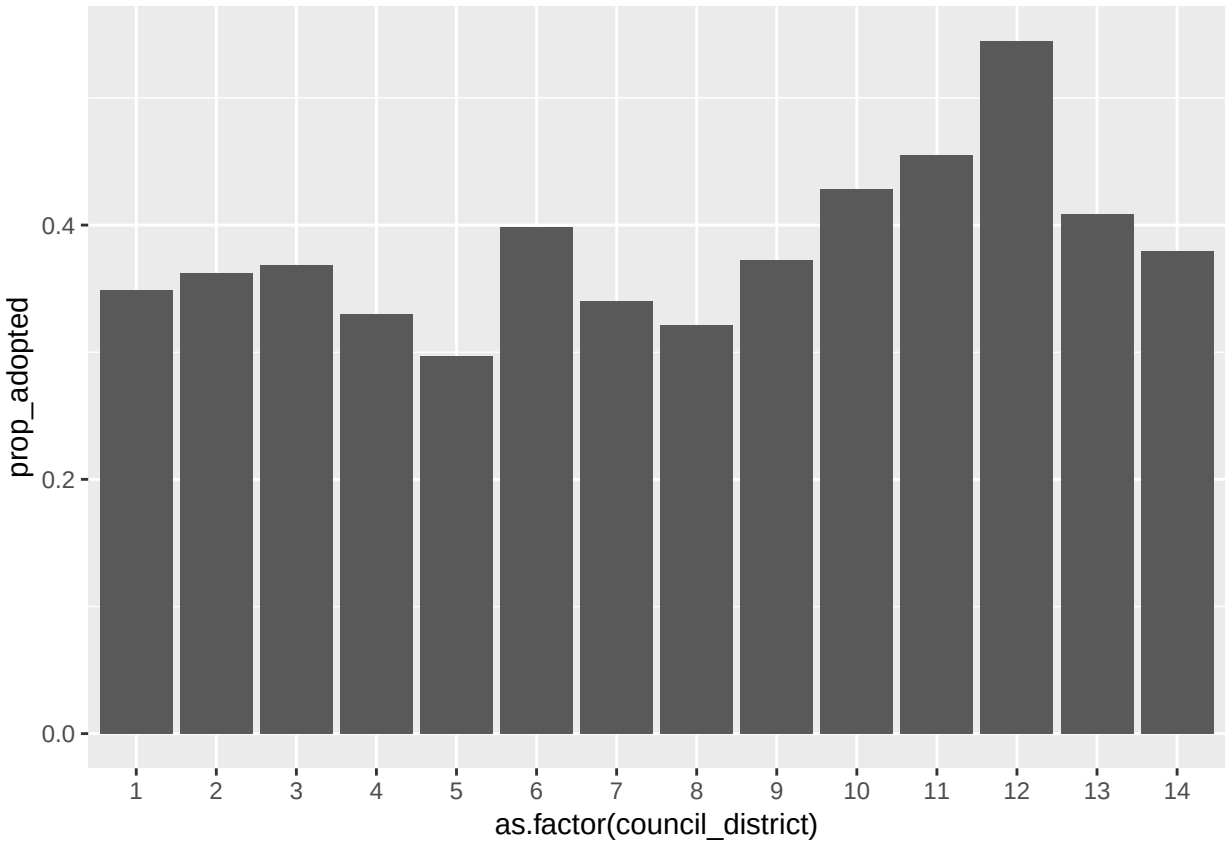
During further EDA we found that there are differences between city council district in the proportion of dogs adopted and dogs that do not leave the shelter system alive. So we tried a multilevel modeling to account for these differences.

```
just_dog_adoptions %>%
  group_by(council_district)%>%
  summarise(prop_dead = sum(out_dead)/n())%>%
  filter(!is.na(council_district))%>%
  ggplot(aes(x=as.factor(council_district), y = prop_dead)) +
  geom_bar(stat = "identity")
```



Above you can see that certain city council districts have higher rates of dogs dying in the shelter system. City council district 12 seems to be the highest at 20.7% of the dogs are dying in the shelter system, and city council district 11 has the lowest at only 13.1% of the dogs dying in the shelter system.

```
just_dog_adoptions %>%
  group_by(council_district)%>%
  summarise(prop_adopted = sum(adopted)/n())%>%
  filter(!is.na(council_district)) %>%
  ggplot(aes(x=as.factor(council_district), y = prop_adopted)) +
  geom_bar(stat = "identity")
```



For proportion adopted we see a range of values. City council district 12 has 54.5% of the dogs adopted, while city council district 5 has the lowest proportion adopted at only 0.297.

First we made a random intercepts model with no level one or level two predictors. While looking at adoptions as our response we found that our one fixed effect $\alpha_0 = -0.4930$, which if you exponentiate and convert to a proportion we see that the average proportion of dogs being adopted is 0.379. It is sad to think that only about one out of every three dogs gets adopted. We fonud a $\sigma_u = 0.2438$ so the average de

Next we repeated the same model with out_dead as the outcome. We fonud a fixed effect for our $\alpha_0 = -1.65465$, which after you exponentiate and convert to a proportion we get 0.160. So only 16% of the dogs that enter the animal shelter system exit dead. We found a $\sigma_u = 0.1395$, so the

```
mod.1=glmer(adopted~1+(1|council_district), data=just_dog_adoptions, family = "binomial")
summary(mod.1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: adopted ~ 1 + (1 | council_district)
##    Data: just_dog_adoptions
##
##      AIC      BIC   logLik deviance df.resid
##  58685.9  58703.3 -29340.9  58681.9    45485
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0569 -0.7335 -0.6885  1.2963  1.5351
##
## Random effects:
```

```
##  Groups           Name        Variance Std.Dev.
##  council_district (Intercept) 0.05945  0.2438
## Number of obs: 45487, groups:  council_district, 14
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.49300    0.06663  -7.399 1.37e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
exp(fixef(mod.1))/(1+exp(fixef(mod.1)))
```

```
## (Intercept)
##   0.3791861
```

```
mod.1_dead=glmer(out_dead~1+(1|council_district), data=just_dog_adoptions, family = "binomial")
summary(mod.1_dead)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: out_dead ~ 1 + (1 | council_district)
##    Data: just_dog_adoptions
##
##      AIC      BIC   logLik deviance df.resid
##  39522.0  39539.5 -19759.0  39518.0    45485
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.4909 -0.4592 -0.4080 -0.4014  2.4913
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  council_district (Intercept) 0.01946  0.1395
## Number of obs: 45487, groups:  council_district, 14
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.65465    0.04143  -39.94   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
exp(fixef(mod.1_dead))/(1+exp(fixef(mod.1_dead)))
```

```
## (Intercept)
##   0.1604818
```

We repeated the same procedure as above just this time adding in summer, chip status and treatable intake. We found with adopted as the response the average proportion adopted is 0.126 not durring the summer for dogs that are untreatable and the chip was not readable or not present. We fonud

```
mod.2.adopted=glmer(adopted~summer+chip_status+treatable_intake+(1|council_district), data=just_dog_adop
summary(mod.2.adopted)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
```

```
## Formula:
## adopted ~ summer + chip_status + treatable_intake + (1 | council_district)
##    Data: just_dog_adoptions
##
##      AIC      BIC   logLik deviance df.resid
##  57316.1  57359.7 -28653.0  57306.1    45482
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.2426 -0.7608 -0.6787  1.2118  3.4802
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  council_district (Intercept) 0.08389  0.2896
## Number of obs: 45487, groups:  council_district, 14
##
## Fixed effects:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.93608    0.09252 -20.926  < 2e-16 ***
## summer            0.07182    0.02272   3.162  0.00157 **
## chip_status      -0.12224    0.02227  -5.490 4.03e-08 ***
## treatable_intake  1.59669    0.05142  31.050  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) summer chp_st
## summer      -0.069
## chip_status -0.073  0.027
## treatbl_ntk -0.519  0.002 -0.014
```

```
exp(fixef(mod.2.adopted))/(1+exp(fixef(mod.2.adopted)))
```

```
##      (Intercept)           summer      chip_status treatable_intake
##        0.1260794        0.5179471        0.4694784        0.8315547
```

```
mod.2.dead=glmer(out_dead~summer+chip_status+treatable_intake+(1|council_district), data=just_dog_adopt
summary(mod.2.dead)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## out_dead ~ summer + chip_status + treatable_intake + (1 | council_district)
##    Data: just_dog_adoptions
##
##      AIC      BIC   logLik deviance df.resid
##  36001.9  36045.5 -17996.0  35991.9    45482
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3865 -0.3979 -0.3527 -0.3135  3.4388
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
```

```
##  council_district (Intercept) 0.01982  0.1408
## Number of obs: 45487, groups:  council_district, 14
##
## Fixed effects:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.08265    0.05258   1.572    0.116
## summer            0.37470    0.02963  12.644  < 2e-16 ***
## chip_status      -0.23570    0.03116  -7.565 3.88e-14 ***
## treatable_intake -2.12055    0.03592 -59.039  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) summer chp_st
## summer      -0.149
## chip_status -0.194  0.020
## treatbl_ntk -0.530 -0.051  0.021
```

```
exp(fixef(mod.2.dead))/(1+exp(fixef(mod.2.dead)))
```

```
##      (Intercept)           summer      chip_status treatable_intake
##        0.5206502        0.5925930        0.4413468        0.1071159
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, May 03, 2019 - 11:43:11 AM

Table 1:

	Dependent variable:	
	out_dead	
	(1)	(2)
summer		0.375***
		(0.030)
chip_status		−0.236***
		(0.031)
treatable_intake		−2.121***
		(0.036)
Constant	−1.655***	0.083
	(0.041)	(0.053)
District σ	0.1395	0.1408
Observations	45,487	45,487
Log Likelihood	−19,759.010	−17,995.950
Akaike Inf. Crit.	39,522.010	36,001.900
Bayesian Inf. Crit.	39,539.460	36,045.530
Note:	*p<0.1; **p<0.05; ***p<0.01	