Case Study 1 - Explore

Weiting Miao

2023-10-16

Questions of Interest

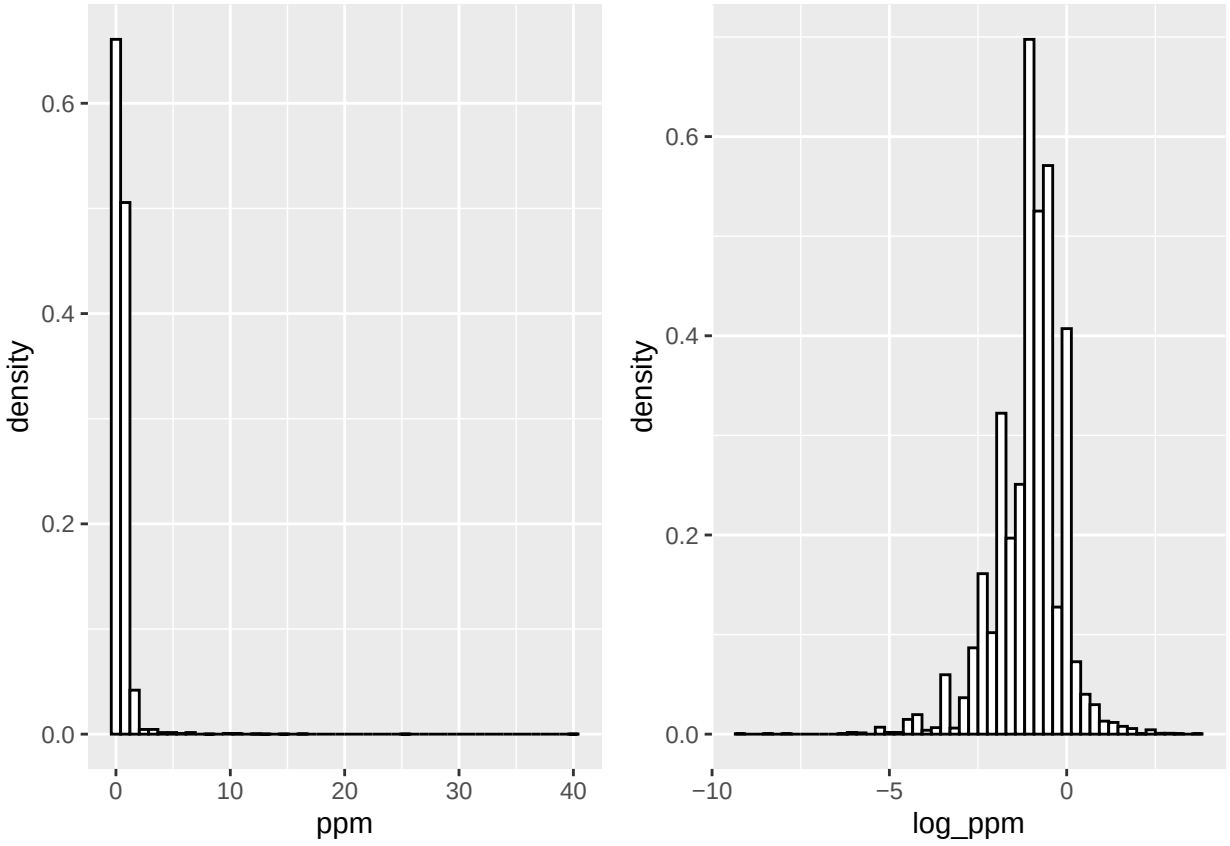
To investigate factors related to the price per mg of your drug, accounting for potential clustering by location and exploring heterogeneity in pricing by location.

Data Cleaning

Variables:

- Outcome variable: ppm
- Purchasing time: year, month, day, quarter
- Location: country, state, city, USA_region - country: all US; city: many missing
- Source: source of information - Unknown
- Formulation of the drug: factor - no variation, drop it
- Dosage strength - whether treat it as factor or numeric variable, or group them
- Bulk Purchase: dummy
- Primary reason: factor - many missing values

(1). Drop observations if the outcome variable is missing



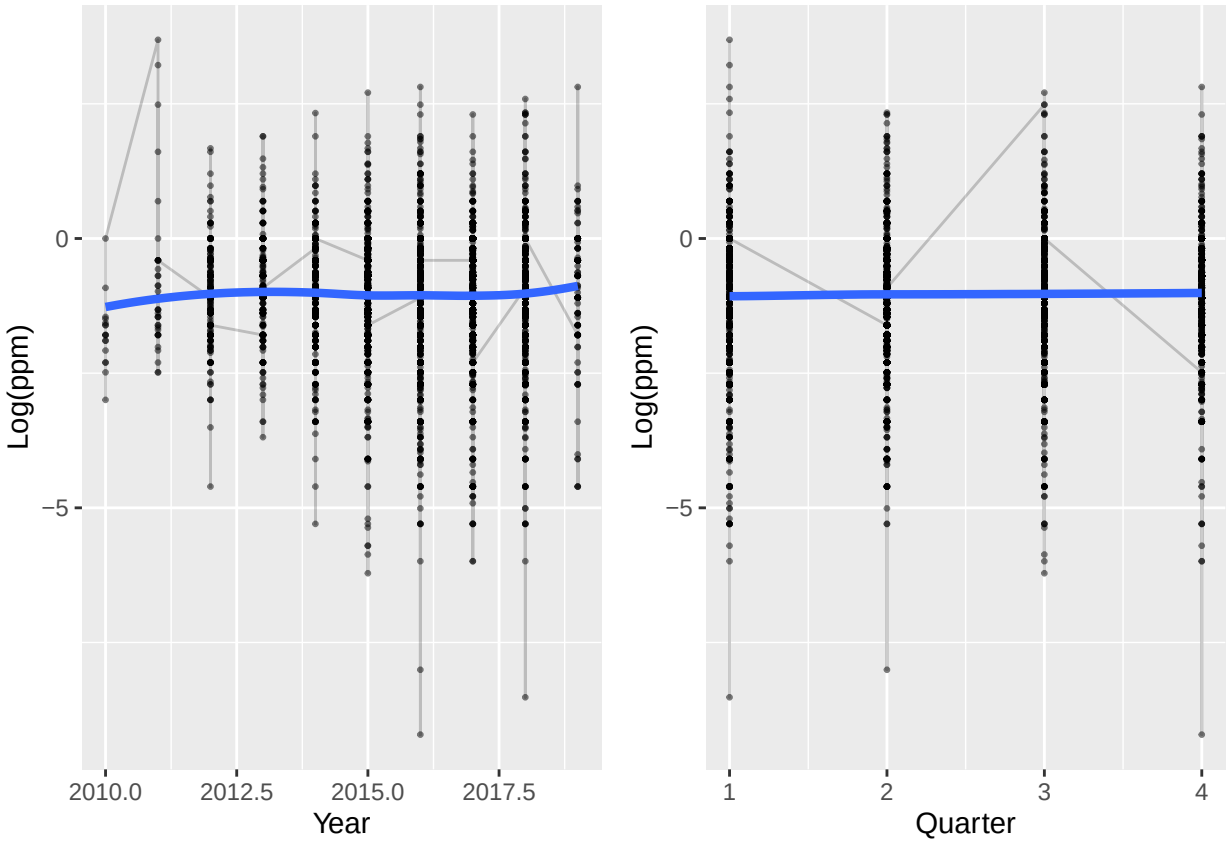
(2). Drop variables without any variation - country, formulation

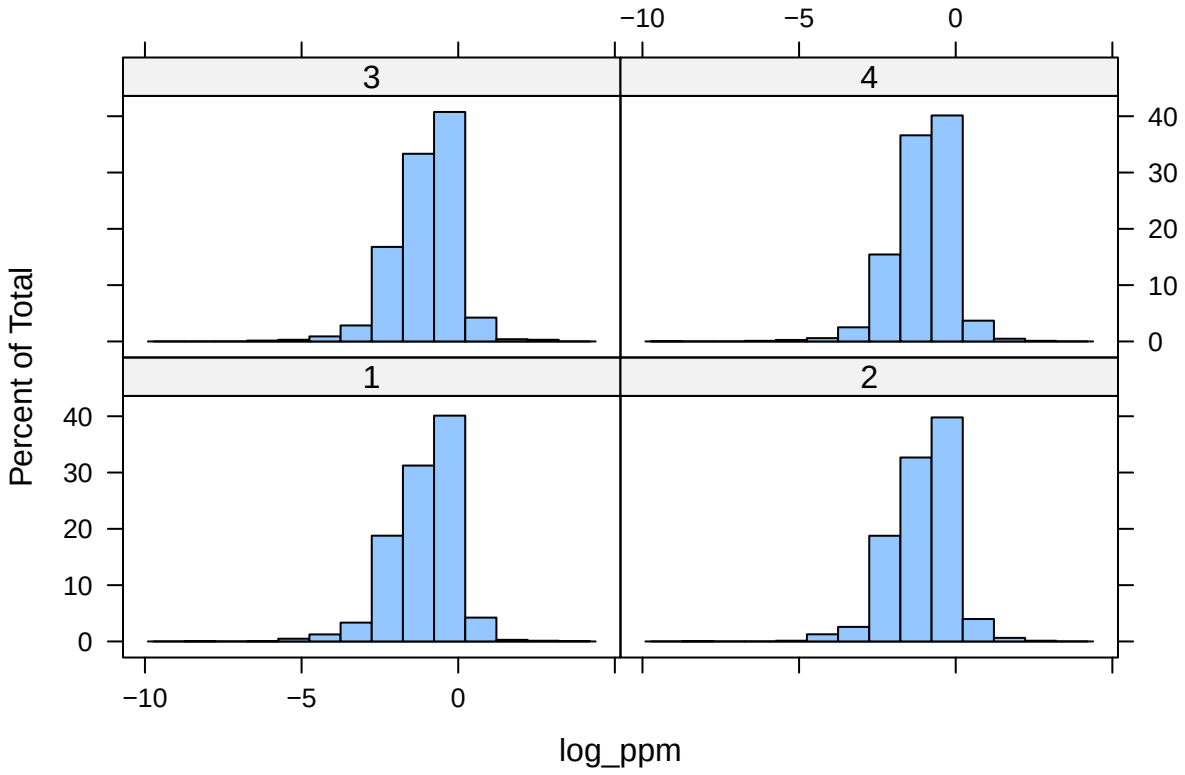
(3). Remove states with 1 observation, otherwise, we cannot estimate the within-group variance for these states. State has 1 missing values, and it has wrong category "USA". Since our goal is to understand the heterogeneity across locations, we remove observations with "USA" as the state name. City has a lot of missing values, so we will focus on the heterogeneity across states.

state	Count
California	1041
Florida	638
Texas	480
Michigan	439
Pennsylvania	377
Arizona	360
Ohio	313
New York	306
Washington	288
Tennessee	253
Georgia	236
Oklahoma	230
Indiana	225
Oregon	224
Illinois	214
Missouri	213
North Carolina	204
Colorado	199
Alabama	187
Nevada	171
Wisconsin	156
South Carolina	136
Virginia	136
Arkansas	135
New Jersey	133
Massachusetts	126
Maryland	119
Minnesota	110
Iowa	106
Kentucky	99
Louisiana	96
Kansas	89
Utah	75
Mississippi	62
Connecticut	60
Nebraska	54
USA	53
New Mexico	48
Hawaii	44
West Virginia	43
Idaho	39
Montana	35
Maine	30
Delaware	26
Rhode Island	20
Alaska	18
New Hampshire	17
North Dakota	13
Wyoming	12
Vermont	10
Washington, DC	10
South Dakota	9
Puerto Rico	2
	1
American Samoa	1
Guam	1
Midway Atoll	1

USA_region	Count
South	3090
West	2554
Midwest	1941
Northeast	1079
Other/Unknown	55

(4). Create year and quarter variables. Very limited observations before year 2010. Given the nonlinear effects of year and quarter, we consider to treat year and quarter as factor variables.

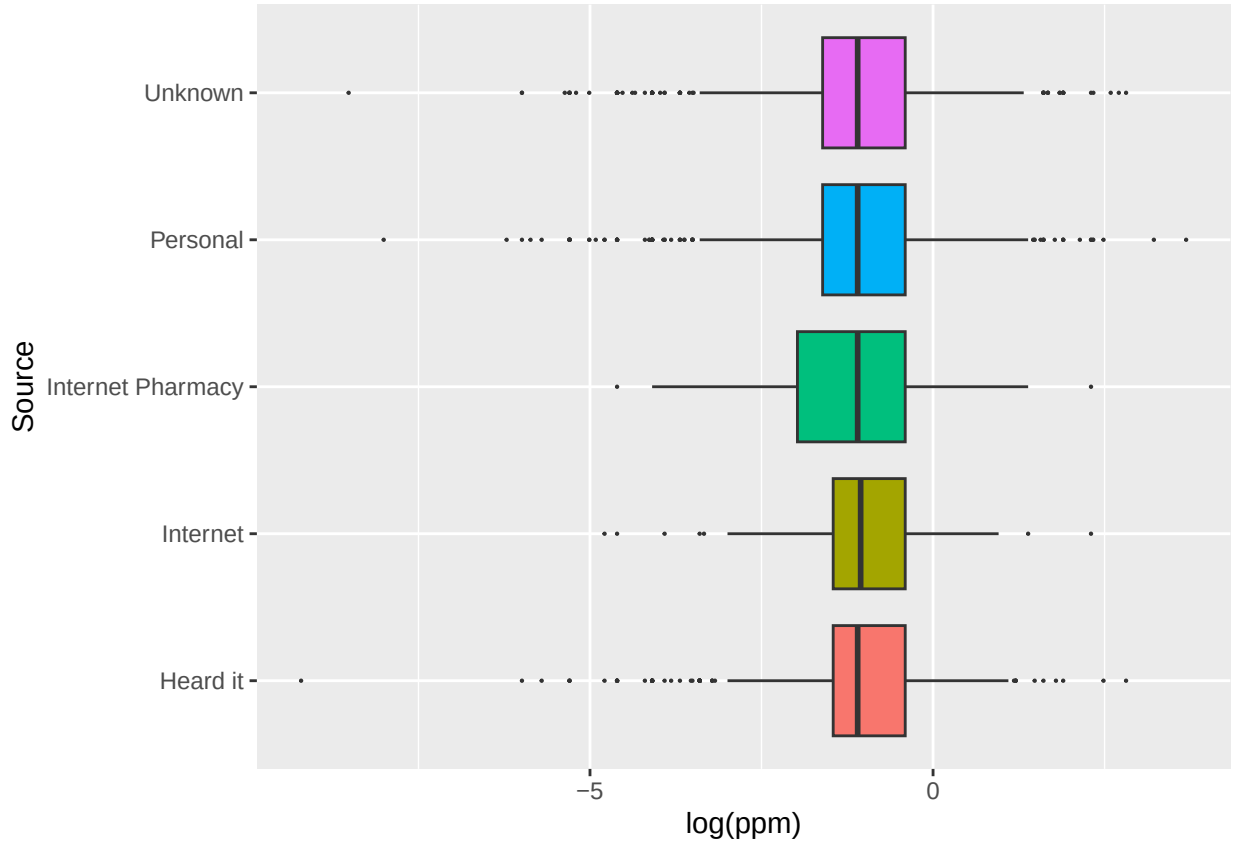




Quarter doesn't seem to be important.

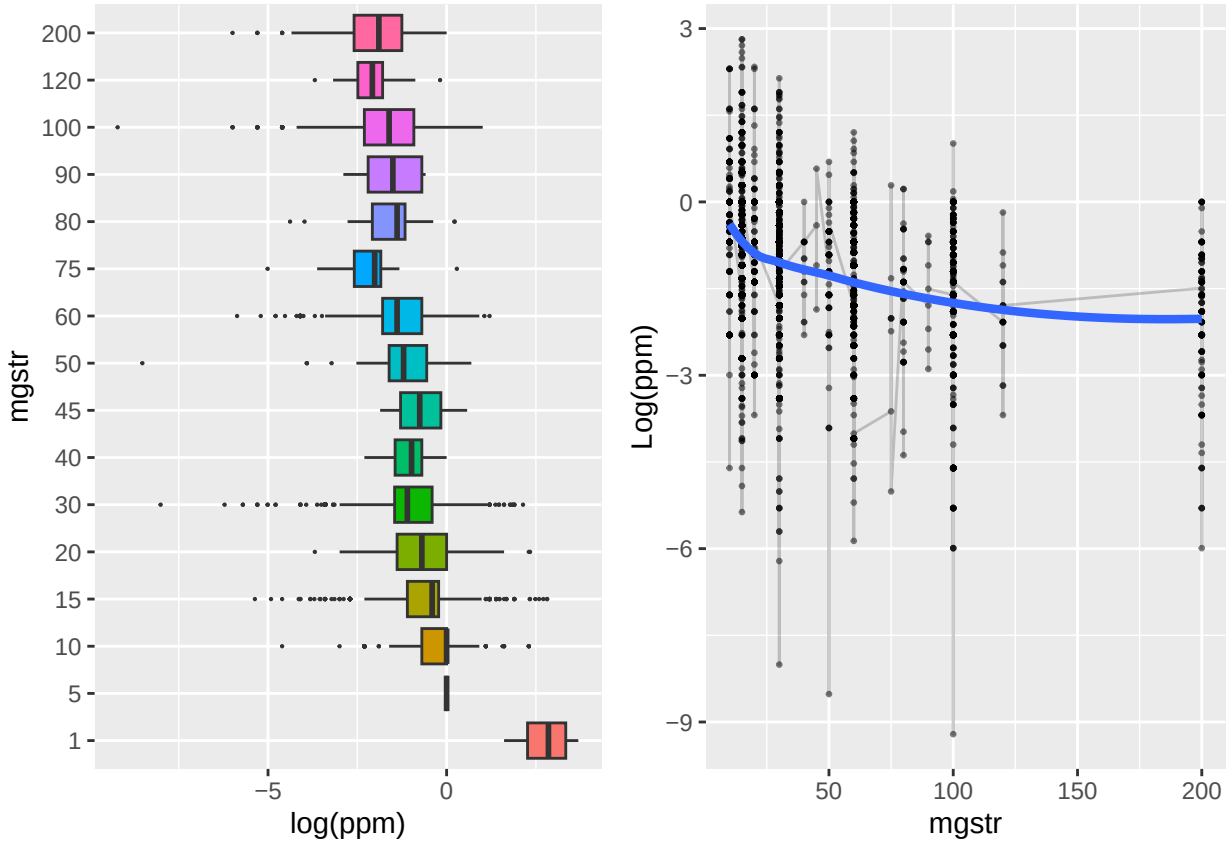
(5). Source: from the raw data, most can be coded as "Internet".

source
Personal
Heard it
Internet
Internet Pharmacy
http://forum.opiophile.org/showthread.php?37268-NationWide-OC-amp-Hydro-Morphone-Survey-Important-*Personal-Re
http://www.bluelight.ru/vb/showthread.php?t=480752&page=9
http://silkroadvb5piz3r.onion/silkroad/category/240/50
http://silkroadvb5piz3r.onion/silkroad/category/51/200
http://silkroadvb5piz3r.onion/silkroad/category/51/25
http://silkroadvb5piz3r.onion/silkroad/category/51
http://wiki.answers.com/Q/What_is_the_street_cost_of_morphine_sulfate_100mg_a_pill&dim1=UpdatedAns&dim2=
http://www.bluelight.ru/vb/showthread.php?t=480752&page=2
http://silkroadvb5piz3r.onion/silkroad/category/51/150
http://silkroadvb5piz3r.onion/silkroad/category/51/50
http://www.bluelight.ru/vb/showthread.php?t=480752&page=3
http://www.bluelight.ru/vb/showthread.php?t=480752&page=4
Bluelight.com
http://silkroadvb5piz3r.onion/silkroad/category/51/175
https://www.drugs.com/price-guide/morphine
Drug forum
Drug forum.com
Fuckoff.org
Google
google.com
http://silkroadvb5piz3r.onion/index.php/silkroad/item/06ee6c0cfb
http://silkroadvb5piz3r.onion/index.php/silkroad/item/268af26749
http://silkroadvb5piz3r.onion/index.php/silkroad/item/2db123a294
http://silkroadvb5piz3r.onion/index.php/silkroad/item/35e2f9fff0
http://silkroadvb5piz3r.onion/index.php/silkroad/item/4603fc6c4c
http://silkroadvb5piz3r.onion/index.php/silkroad/item/5442e57d37
http://silkroadvb5piz3r.onion/index.php/silkroad/item/7259fa0fe6
http://silkroadvb5piz3r.onion/index.php/silkroad/item/80acbc7f83
http://silkroadvb5piz3r.onion/index.php/silkroad/item/8ba6726a41
http://silkroadvb5piz3r.onion/index.php/silkroad/item/8e3b43f251
http://silkroadvb5piz3r.onion/index.php/silkroad/item/9dba78d600
http://silkroadvb5piz3r.onion/index.php/silkroad/item/a74ce13365
http://silkroadvb5piz3r.onion/index.php/silkroad/item/a97c7d2ac8
http://silkroadvb5piz3r.onion/index.php/silkroad/item/bcdebee234
http://silkroadvb5piz3r.onion/index.php/silkroad/item/f5790bab0e
http://silkroadvb5piz3r.onion/index.php/silkroad/item/f7774a3fca
http://silkroadvb5piz3r.onion/silkroad/category/51/125
http://silkroadvb5piz3r.onion/silkroad/category/51/225
http://silkroadvb5piz3r.onion/silkroad/category/51/75
http://silkroadvb5piz3r.onion/silkroad/item/03adfe43ed
http://silkroadvb5piz3r.onion/silkroad/item/56fe5ee7bb
http://silkroadvb5piz3r.onion/silkroad/item/c452141bb3
http://silkroadvb5piz3r.onion/silkroad/item/cf888a9253
L
Reddit.com
USAsomething.com



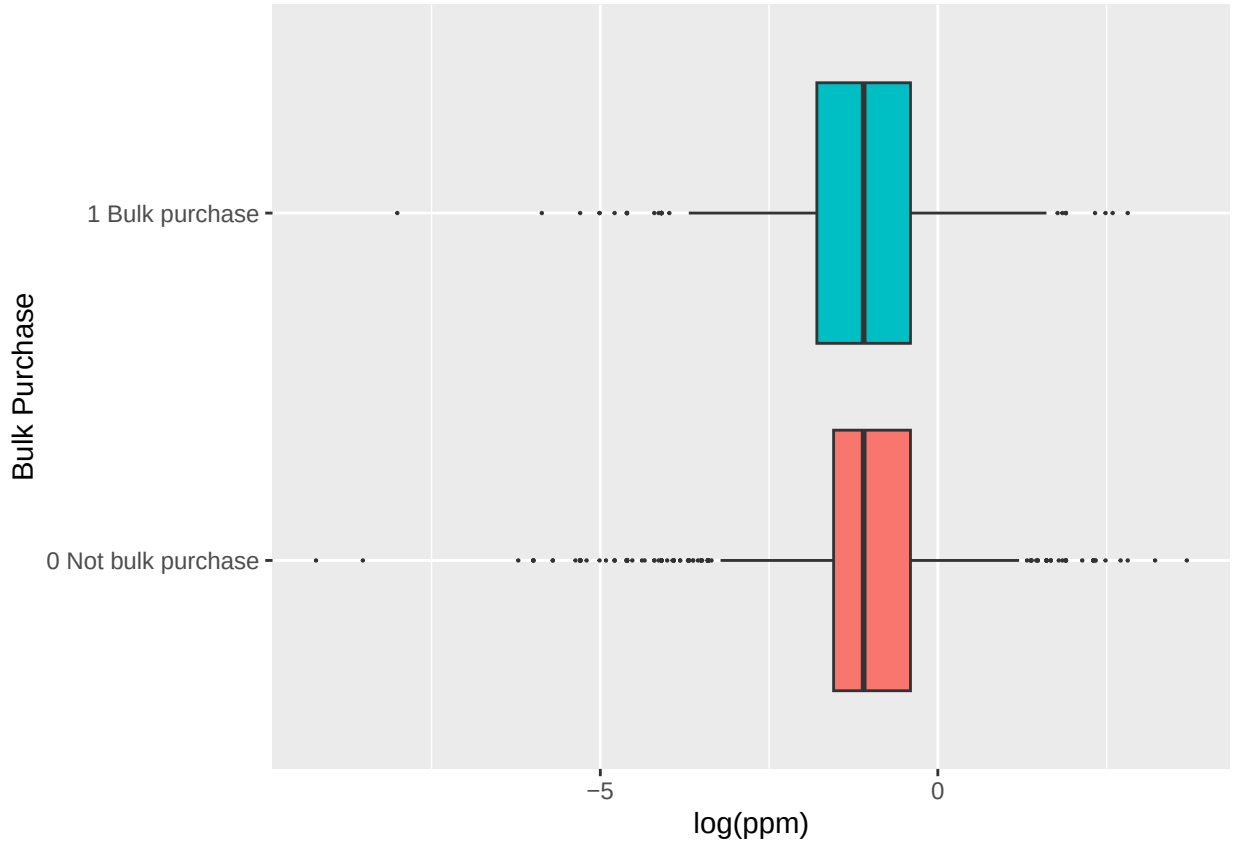
(6). Dosage strength: should we group them?

```
##
##    1    5   10   15   20   30   40   45   50   60   75   80   90  100  120  200
##    4    1  286 2559  205 3399   16    4   78 1246    8   47    9  697   21  139
```



The effects are mostly monotone and linear except for when mgstr < 10 (5 obs).

(7). Bulk Purchase



(8). Primary reason: 70% of the observations are NA or didn't answer the question. We decide not to use this variable.

Primary_Reason	Count
	4750
0 Reporter did not answer the question	1424
9 To self-treat my pain	1041
8 Prefer not to answer	805
6 Other reason	170
4 For enjoyment/to get high	148
7 Don't know	145
10 To treat a medical condition other than pain	77
5 To resell	76
3 To prevent or treat withdrawal	75
11 To come down	8

9)

```
##
##           film       lollipop lozenge/troche       multiple            N/A
##              0              0              0              0              0
##    nasal spray     oral spray          patch    pill/tablet         powder
##              0              0              0           8714              0
##    suppository   syrup/liquid
##              5              0
```

Model Building & Model Selection

Predictors: year, quarter, mgstr, bulk_purchase, source Random intercept: state Random slope:

(1) Random effect or fixed effect?

```
## [1] 23594.12
```

```
## [1] 23297.73
```

Random effect model has much smaller BIC

(2) Year as numeric variable or factor variable?

```
## Data: df %>% filter(year > 2009 & state != "Unknown")
## Models:
## m2: log_ppm ~ year + quarter_factor + mgstr + bulk_purchase + source + (1 | state)
## m3: log_ppm ~ year_factor + quarter_factor + mgstr + bulk_purchase + source + (1 | state)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m2   22 22947 23102 -11452    22903
## m3   30 22938 23150 -11439    22878 25.218  8   0.001428 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Treat it as numerical variable

(3) mgstr as numeric variable or factor variable?

```
## Data: df %>% filter(year > 2009 & state != "Unknown")
## Models:
## m2: log_ppm ~ year + quarter_factor + mgstr + bulk_purchase + source + (1 | state)
## m4: log_ppm ~ year + quarter_factor + mgstr_factor + bulk_purchase + source + (1 | state)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m2   22 22947 23102 -11452    22903
## m4   36 22683 22937 -11305    22611 292.28 14  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Treat mgstr as factor variable

(4) Do we need quarter?

```
## Data: df %>% filter(year > 2009 & state != "Unknown")
## Models:
## m5: log_ppm ~ year + mgstr_factor + bulk_purchase + source + (1 | state)
## m4: log_ppm ~ year + quarter_factor + mgstr_factor + bulk_purchase + source + (1 | state)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m5   33 22681 22914 -11308    22615
## m4   36 22683 22937 -11305    22611 4.3555  3     0.2255
```

Quarter doesn't seem to be important. Drop quarter.

(5) Do we need year?

```
## Data: df %>% filter(year > 2009 & state != "Unknown")
## Models:
## m6: log_ppm ~ mgstr_factor + bulk_purchase + source + (1 | state)
## m5: log_ppm ~ year + mgstr_factor + bulk_purchase + source + (1 | state)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m6   32 22683 22909 -11310    22619
## m5   33 22681 22914 -11308    22615 3.9088  1    0.04803 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Year doesn't seem to be important. Drop year.

(6) Do we need source?

```
## Data: df %>% filter(year > 2009 & state != "Unknown")
## Models:
## m7: log_ppm ~ mgstr_factor + bulk_purchase + (1 | state)
## m6: log_ppm ~ mgstr_factor + bulk_purchase + source + (1 | state)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m7   19 22681 22815 -11322    22643
## m6   32 22683 22909 -11310    22619 24.118 13    0.03006 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adding source into the regression increases BIC, get rid of source

(7) Any random slope

```
## Data: df %>% filter(year > 2009 & state != "Unknown")
## Models:
## m7: log_ppm ~ mgstr_factor + bulk_purchase + (1 | state)
## m8: log_ppm ~ mgstr_factor + bulk_purchase + (1 + bulk_purchase | state)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m7   19 22681 22815 -11322    22643
## m8   21 22683 22831 -11320    22641 2.3343  2     0.3113
```

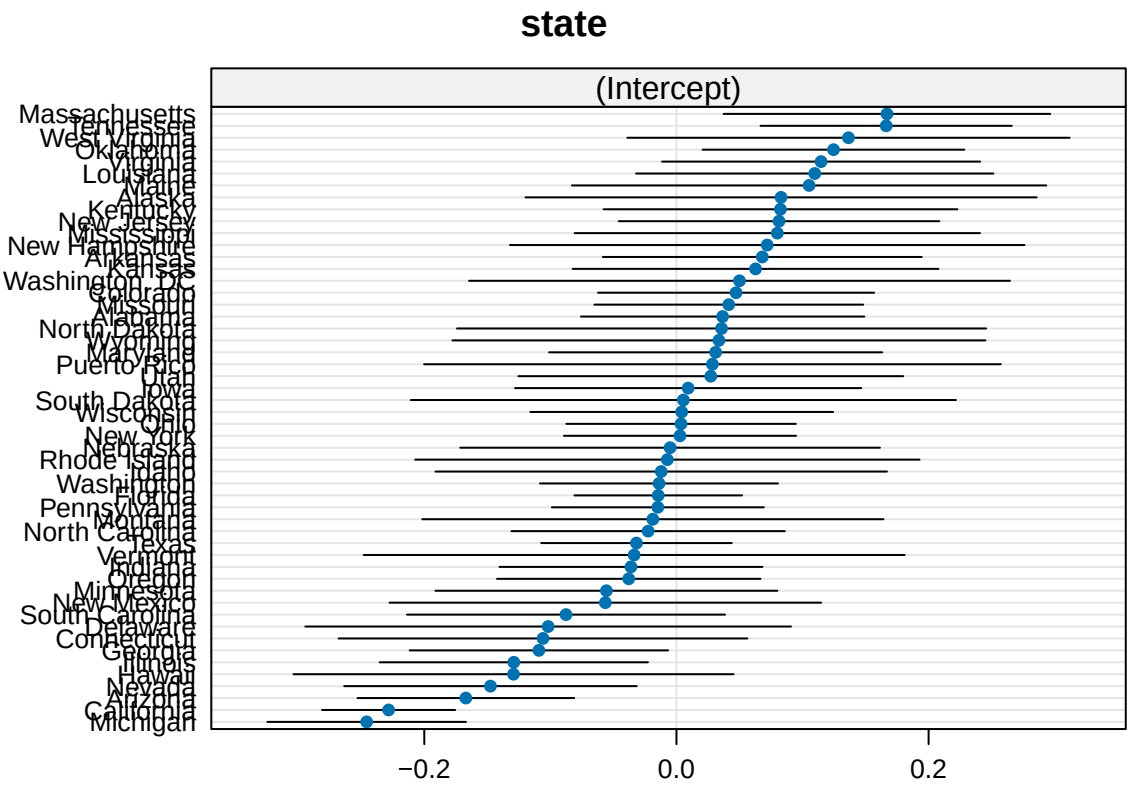
Having random slope of bulk purchase doesn't improve the performance

(8) Any interaction term

```
## Data: df %>% filter(year > 2009 & state != "Unknown")
## Models:
## m7: log_ppm ~ mgstr_factor + bulk_purchase + (1 | state)
## m9: log_ppm ~ mgstr_factor * bulk_purchase + (1 | state)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m7   19 22681 22815 -11322    22643
## m9   31 22676 22895 -11307    22614 28.935 12   0.004028 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

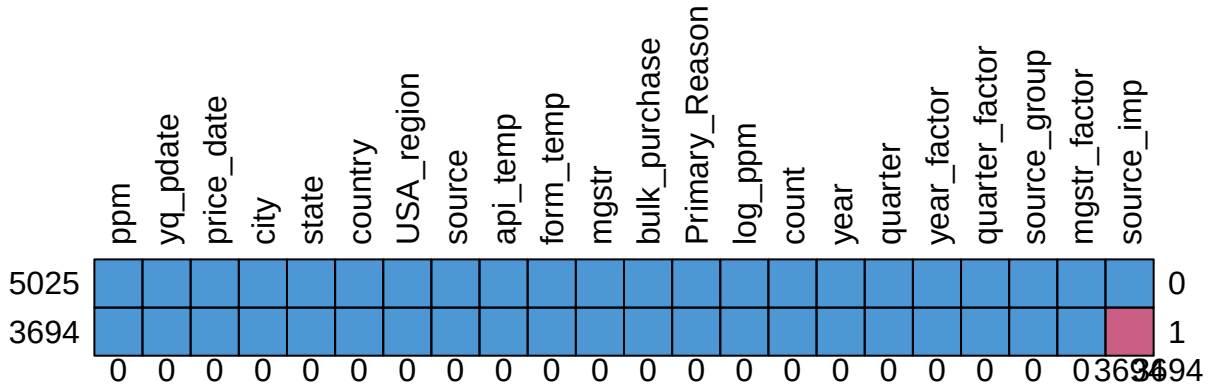
Not helpful to include the interaction term.

The final model is m7.



Missing data imputation

Even though we choose to drop source variable, we decide to check if source useful after we impute missing data of source. We will use MICE to do missing data imputation. Let's first look into EDA of missing values.
preprocess data




```
##      ppm yq_pdate price_date city state country USA_region source api_temp
## 5025   1        1          1    1     1       1          1      1        1
## 3694   1        1          1    1     1       1          1      1        1
##        0        0          0    0     0       0          0      0        0
##      form_temp mgstr bulk_purchase Primary_Reason log_ppm count year quarter
## 5025         1     1             1              1       1     1    1       1
## 3694         1     1             1              1       1     1    1       1
##              0     0             0              0       0     0    0       0
##      year_factor quarter_factor source_group mgstr_factor source_imp
## 5025           1              1            1            1          1    0
## 3694           1              1            1            1          0    1
##                0              0            0            0       3694 3694
```

At the bottom: total number of missing values by variables. On the right: number of variables missing in each pattern. On the left: number of cases for each pattern.

Missing data: 1)State has 1 missing values, and it has wrong category "USA", so we also map to unknown 2)City has 3126 missing values and duplicate entries as well as abbreviations, MICE couldn't handle variable with over 50 catogories, so we drop city. 3)source has 3695 missing values and messy website. 4)Primary_Reason has 4754 missing values # drop it because of over 60% missing rate

source_imp
NA
Personal
Heard it
Internet
Internet Pharmacy
http://forum.opiophile.org/showthread.php?37268-NationWide-OC-amp-Hydro-Morphone-Survey-Important-*Personal-Re
http://www.bluelight.ru/vb/showthread.php?t=480752&page=9
http://silkroadvb5piz3r.onion/silkroad/category/240/50
http://silkroadvb5piz3r.onion/silkroad/category/51/200
http://silkroadvb5piz3r.onion/silkroad/category/51/25
http://silkroadvb5piz3r.onion/silkroad/category/51
http://wiki.answers.com/Q/What_is_the_street_cost_of_morphine_sulfate_100mg_a_pill&dim1=UpdatedAns&dim2=
http://www.bluelight.ru/vb/showthread.php?t=480752&page=2
http://silkroadvb5piz3r.onion/silkroad/category/51/150
http://silkroadvb5piz3r.onion/silkroad/category/51/50
http://www.bluelight.ru/vb/showthread.php?t=480752&page=3
http://www.bluelight.ru/vb/showthread.php?t=480752&page=4
Bluelight.com
http://silkroadvb5piz3r.onion/silkroad/category/51/175
https://www.drugs.com/price-guide/morphine
Drug forum
Drug forum.com
Fuckoff.org
Google
google.com
http://silkroadvb5piz3r.onion/index.php/silkroad/item/06ee6c0cfb
http://silkroadvb5piz3r.onion/index.php/silkroad/item/268af26749
http://silkroadvb5piz3r.onion/index.php/silkroad/item/2db123a294
http://silkroadvb5piz3r.onion/index.php/silkroad/item/35e2f9fff0
http://silkroadvb5piz3r.onion/index.php/silkroad/item/4603fc6c4c
http://silkroadvb5piz3r.onion/index.php/silkroad/item/5442e57d37
http://silkroadvb5piz3r.onion/index.php/silkroad/item/7259fa0fe6
http://silkroadvb5piz3r.onion/index.php/silkroad/item/80acbc7f83
http://silkroadvb5piz3r.onion/index.php/silkroad/item/8ba6726a41
http://silkroadvb5piz3r.onion/index.php/silkroad/item/8e3b43f251
http://silkroadvb5piz3r.onion/index.php/silkroad/item/9dba78d600
http://silkroadvb5piz3r.onion/index.php/silkroad/item/a74ce13365
http://silkroadvb5piz3r.onion/index.php/silkroad/item/a97c7d2ac8
http://silkroadvb5piz3r.onion/index.php/silkroad/item/bcdebee234
http://silkroadvb5piz3r.onion/index.php/silkroad/item/f5790bab0e
http://silkroadvb5piz3r.onion/index.php/silkroad/item/f7774a3fca
http://silkroadvb5piz3r.onion/silkroad/category/51/125
http://silkroadvb5piz3r.onion/silkroad/category/51/225
http://silkroadvb5piz3r.onion/silkroad/category/51/75
http://silkroadvb5piz3r.onion/silkroad/item/03adfe43ed
http://silkroadvb5piz3r.onion/silkroad/item/56fe5ee7bb
http://silkroadvb5piz3r.onion/silkroad/item/c452141bb3
http://silkroadvb5piz3r.onion/silkroad/item/cf888a9253
L
Reddit.com
USAsomething.com

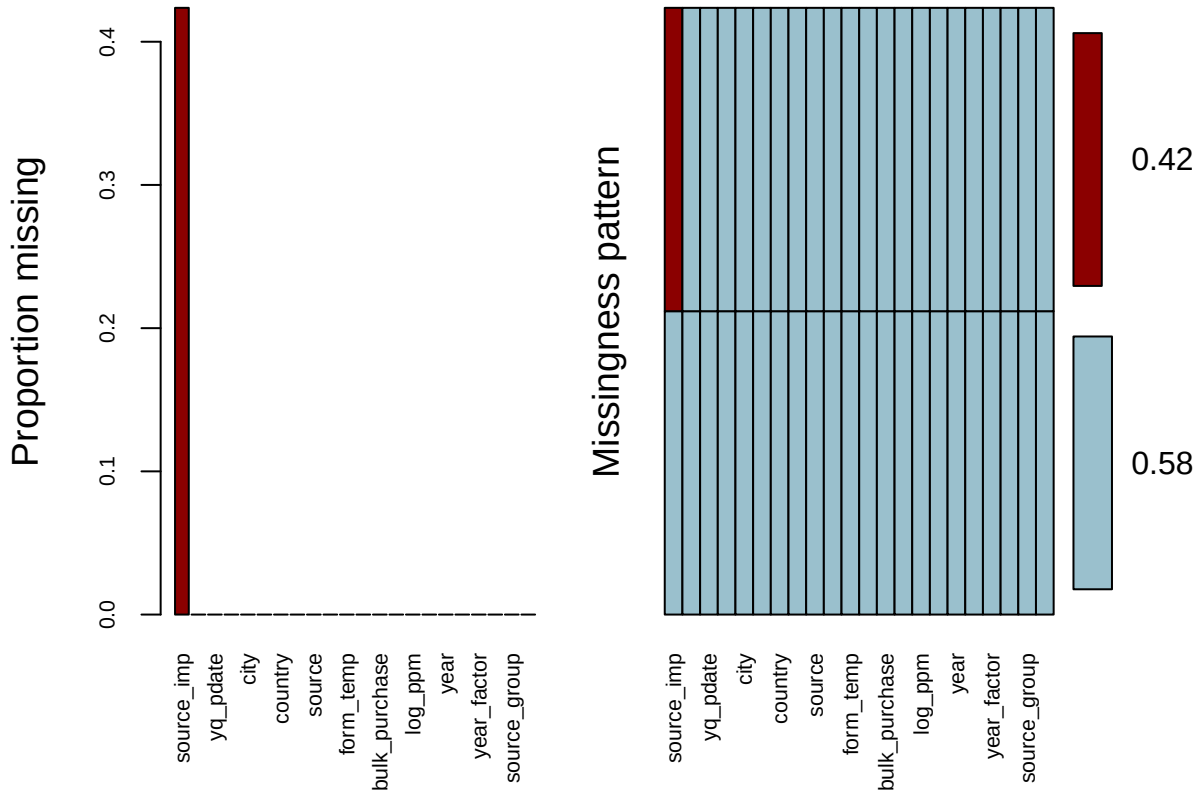
We need to group source_imp.

source_imp	Count
NA	3694
Personal	2941
Heard it	1621
Internet	353
Internet Pharmacy	110

Imputation

We assume that the data is MAR, the next step is to use MICE to impute missing data.

One problem: city has over 50 categories and is messed up with duplicate and abbreviation, so here we will drop city.




```
##
##  Variables sorted by number of missings:
##        Variable     Count
##      source_imp 0.4236724
##             ppm 0.0000000
##        yq_pdate 0.0000000
##      price_date 0.0000000
##            city 0.0000000
##           state 0.0000000
```

```
##         country 0.0000000
##      USA_region 0.0000000
##          source 0.0000000
##        api_temp 0.0000000
##       form_temp 0.0000000
##           mgstr 0.0000000
##   bulk_purchase 0.0000000
##  Primary_Reason 0.0000000
##         log_ppm 0.0000000
##           count 0.0000000
##            year 0.0000000
##         quarter 0.0000000
##     year_factor 0.0000000
##  quarter_factor 0.0000000
##    source_group 0.0000000
##    mgstr_factor 0.0000000
```

The typical sequence of steps to perform a multiple imputation analysis is:

1) Impute the missing data by the mice() function, resulting in a multiple imputed data set (class mids);
2) Fit the model of interest (scientific model) on each imputed data set by the with() function, resulting an object of class mira;
3) Pool the estimates from each model into a single set of estimates and standard errors, resulting in an object of class mipo;
4) Optionally, compare pooled estimates from different scientific models by the D1() or D3() functions.

```
##
##  iter imp variable
##   1   1  source_imp
##   1   2  source_imp
##   1   3  source_imp
##   1   4  source_imp
##   1   5  source_imp
##   2   1  source_imp
##   2   2  source_imp
##   2   3  source_imp
##   2   4  source_imp
##   2   5  source_imp
##   3   1  source_imp
##   3   2  source_imp
##   3   3  source_imp
##   3   4  source_imp
##   3   5  source_imp
##   4   1  source_imp
##   4   2  source_imp
##   4   3  source_imp
##   4   4  source_imp
##   4   5  source_imp
##   5   1  source_imp
##   5   2  source_imp
##   5   3  source_imp
##   5   4  source_imp
##   5   5  source_imp
```

model selection

The standard multiple imputation scheme consists of three phases:

1. Imputation of the missing data m = 5 times;
2. Analysis of the m = 5 imputed datasets;
3. Pooling of the parameters across m = 5 analyses.

The final model we have selected is m7 <- lmer(formula = log_ppm ~ mgstr_factor + bulk_purchase + (1 | state), data=df %>% filter(year>2009 & state!="Unknown")) m7 BIC is 22815.

We compare m7 with models that include the source imputed by MICE.

```
## # A tibble: 10 x 10
##    term       npar    AIC    BIC  logLik deviance statistic    df  p.value  nobs
##    <chr>     <dbl>  <dbl>  <dbl>   <dbl>    <dbl>     <dbl> <dbl>    <dbl> <int>
##  1 lmer(log~    19 22681. 22815. -11322.   22643.      NA      NA NA           0
##  2 lmer(log~    22 22661. 22817. -11309.   22617.      25.7     3  1.11e-5     0
##  3 lmer(log~    19 22681. 22815. -11322.   22643.      NA      NA NA           0
##  4 lmer(log~    22 22668. 22823. -11312.   22624.      19.3     3  2.34e-4     0
##  5 lmer(log~    19 22681. 22815. -11322.   22643.      NA      NA NA           0
##  6 lmer(log~    22 22658. 22813. -11307.   22614.      29.2     3  2.08e-6     0
##  7 lmer(log~    19 22681. 22815. -11322.   22643.      NA      NA NA           0
##  8 lmer(log~    22 22667. 22822. -11311.   22623.      20.5     3  1.32e-4     0
##  9 lmer(log~    19 22681. 22815. -11322.   22643.      NA      NA NA           0
## 10 lmer(log~    22 22649. 22805. -11303.   22605.      37.7     3  3.28e-8     0
```

```
## [1] 22815.36
```

```
## [1] 22816.08
```

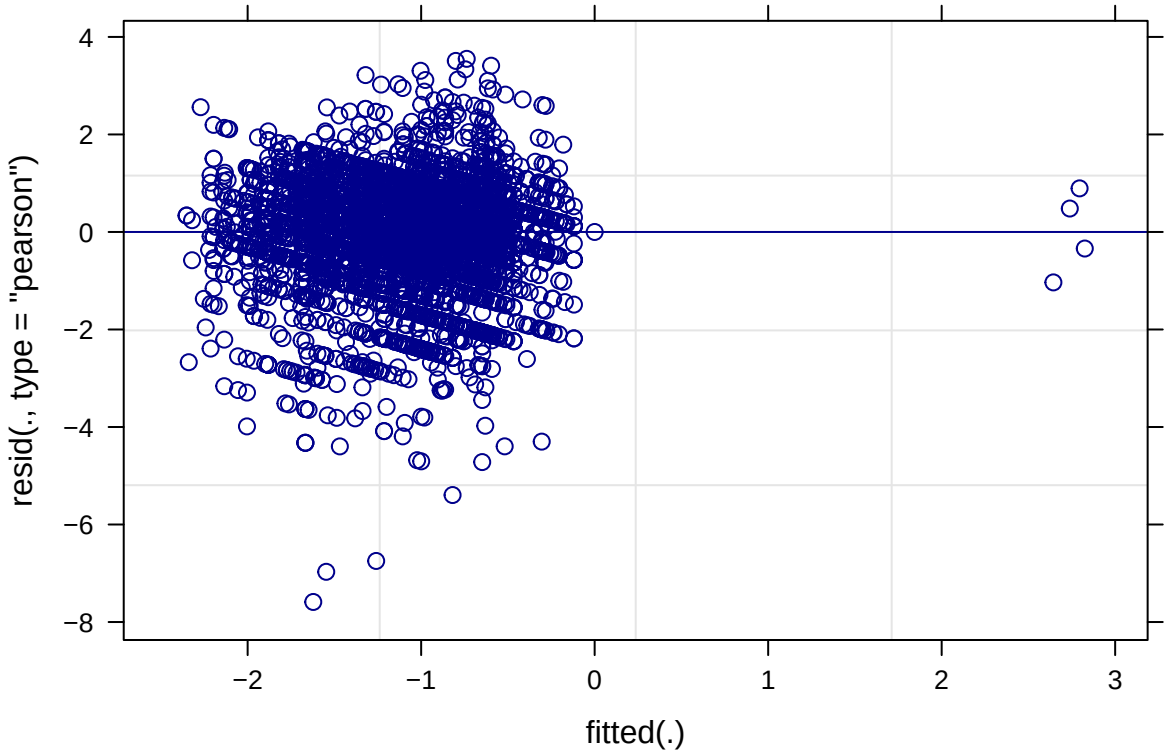
The average BIC on model: lmer(log_ppm ~ mgstr_factor + bulk_purchase + source_imp + (1 | state) is 22816>22815, so the model built with imputed source variable doesn't perform better. We conclude that using m7 model on original data performs better.

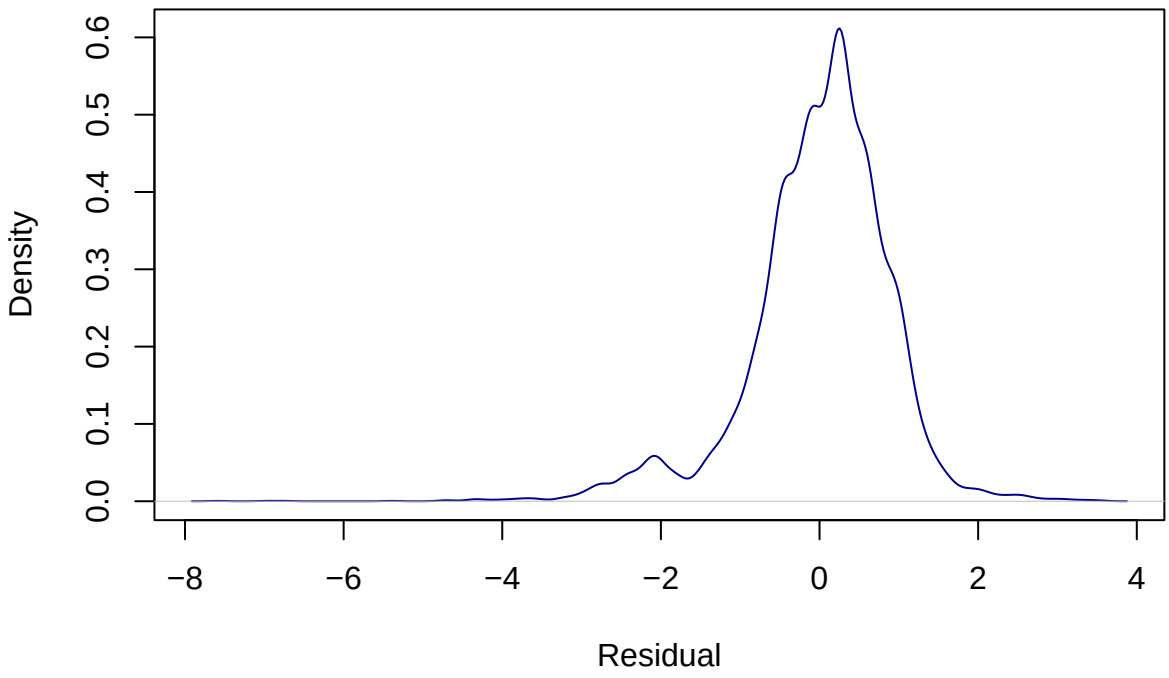
To do:

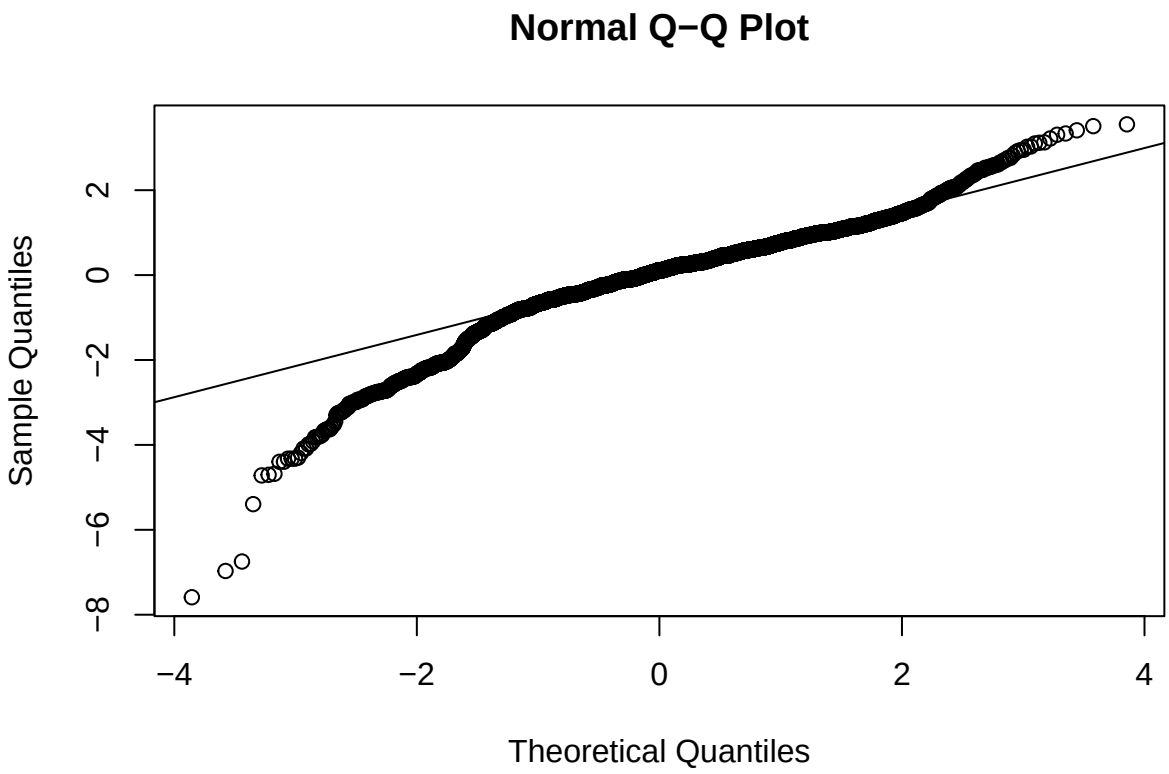
1. Missing data imputation + Adding source into the model and compare BIC(done)
2. Outliers or influential points
3. Model diagnostics
4. Model interpretation + Results visualization

Model Diagnostics

Residual Analysis




Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−8
−6
−4
−2
0
2
4
Residual



Evaluation 3 plots

1. The residual distribution plot and residual vs. fitted values plot tell that the model follows constant variance assumption..
2. In Q-Q plot, we can see that the residuals tend to stray from the line quite a bit near the tails, which could indicate that they're not normally distributed.So the model violate the normality assumption

ICC

Uncertainty quantification around effect estimates of interest

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: log_ppm ~ mgstr_factor + bulk_purchase + (1 | state)
##    Data: df %>% filter(year > 2009 & state != "Unknown")
##
## REML criterion at convergence: 22687.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.4971 -0.4883  0.1267  0.6196  3.9759
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
```

```
##  state    (Intercept) 0.0141   0.1187
##  Residual             0.7976   0.8931
## Number of obs: 8654, groups:  state, 52
##
## Fixed effects:
##                                Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)                    -0.28631    0.05688 1953.48745  -5.033 5.26e-07
## mgstr_factor15                 -0.34702    0.05648 8630.18457  -6.144 8.38e-10
## mgstr_factor20                 -0.47124    0.08252 8625.37630  -5.710 1.16e-08
## mgstr_factor30                 -0.69897    0.05579 8627.52523 -12.530  < 2e-16
## mgstr_factor50                 -1.01511    0.11528 8624.13244  -8.806  < 2e-16
## mgstr_factor60                 -1.03761    0.05939 8628.94224 -17.470  < 2e-16
## mgstr_factor80                 -1.29660    0.14113 8617.34133  -9.187  < 2e-16
## mgstr_factor100                -1.36658    0.06365 8631.35129 -21.472  < 2e-16
## mgstr_factor200                -1.68155    0.09495 8631.64396 -17.709  < 2e-16
## mgstr_factor40                 -0.80751    0.22990 8608.25147  -3.512 0.000446
## mgstr_factor75                 -1.83720    0.32065 8606.74852  -5.730 1.04e-08
## mgstr_factor90                 -1.22988    0.30298 8611.77494  -4.059 4.97e-05
## mgstr_factor120                -1.62570    0.20239 8611.07851  -8.032 1.08e-15
## mgstr_factor1                   2.94474    0.45103 8634.01056   6.529 7.00e-11
## mgstr_factor45                 -0.29653    0.45055 8612.83501  -0.658 0.510457
## mgstr_factor5                   0.31803    0.89545 8599.00139   0.355 0.722478
## bulk_purchase1 Bulk purchase   -0.10685    0.02559 8629.46162  -4.176 3.00e-05
##
## (Intercept)                   ***
## mgstr_factor15                ***
## mgstr_factor20                ***
## mgstr_factor30                ***
## mgstr_factor50                ***
## mgstr_factor60                ***
## mgstr_factor80                ***
## mgstr_factor100               ***
## mgstr_factor200               ***
## mgstr_factor40                ***
## mgstr_factor75                ***
## mgstr_factor90                ***
## mgstr_factor120               ***
## mgstr_factor1                 ***
## mgstr_factor45
## mgstr_factor5
## bulk_purchase1 Bulk purchase ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Evaluation

1. The estimated standard error $\sigma = 0.8931$ describes the within-state or remaining unexplained variation.
2. The estimated $\tau = 0.1187$ describes the across-state variation attributed to the random intercept.
3. We can see that the variation between state is small compared to the variation within state.

$$\tau^2 = 0.01408, \sigma^2 = 0.79763$$

So, $ICC = \frac{\tau^2}{\sigma^2+\tau^2} = \frac{0.01408}{0.79763+0.01408} = 0.017$ The ICC is 0.017 so that we almost have no correlation between ppm in the same state.

Confidence Interval

	lower bound	upper bound
(Intercept)	-0.3977734	-0.1750560
mgstr_factor15	-0.4576486	-0.2364473
mgstr_factor20	-0.6328723	-0.3096573
mgstr_factor30	-0.8082846	-0.5897797
mgstr_factor50	-1.2407912	-0.7892737
mgstr_factor60	-1.1539539	-0.9213306
mgstr_factor80	-1.5730649	-1.0203020
mgstr_factor100	-1.4912750	-1.2419943
mgstr_factor200	-1.8676987	-1.4957456
mgstr_factor40	-1.2576656	-0.3572172
mgstr_factor75	-2.4656391	-1.2096346
mgstr_factor90	-1.8231849	-0.6365000
mgstr_factor120	-2.0221433	-1.2294230
mgstr_factor1	2.0623634	3.8294341
mgstr_factor45	-1.1795426	0.5852616
mgstr_factor5	-1.4357166	2.0715248
bulk_purchase1 Bulk purchase	-0.1570025	-0.0567710

The confidence interval for fixed effect is listed above