United States Home Health Care Analysis

Yinfeng Zhou

2020/11/26

Introduction

This project is meant to implement data analysis on the United States Home Health Care Dataset. The motivation of this analysis is to help understanding what factors contributes to a better home health care in rating, and thus help getting knowledge on how we should improve home health care and what the future direction will be in this field. The dataset used in this report is downloaded from Data.Medicare.Gov.

The dataset includes a number of variables that provides measurements of the quality of the home care services. Among the variables that are directly related to the home care service, there are 6 binary variables and 17 continuous variables. The binary varibales anwer those "Yes or No" questions: `Offers Nursing Care Services`, `Offers Physical Therapy Services`, `Offers Occupational Therapy Services`, `Offers Speech Pathology Services`, `Offers Medical Social Services`, `Offers Home Health Aide Services`. The continuous varibales mainly answer those "How often" questions, using the measure percentage as reported. For example, in the variable `How often the home health team began their patients' care in a timely manner`, the number indicates the percentage of the home health team having begun their patients' care in a timely manner.

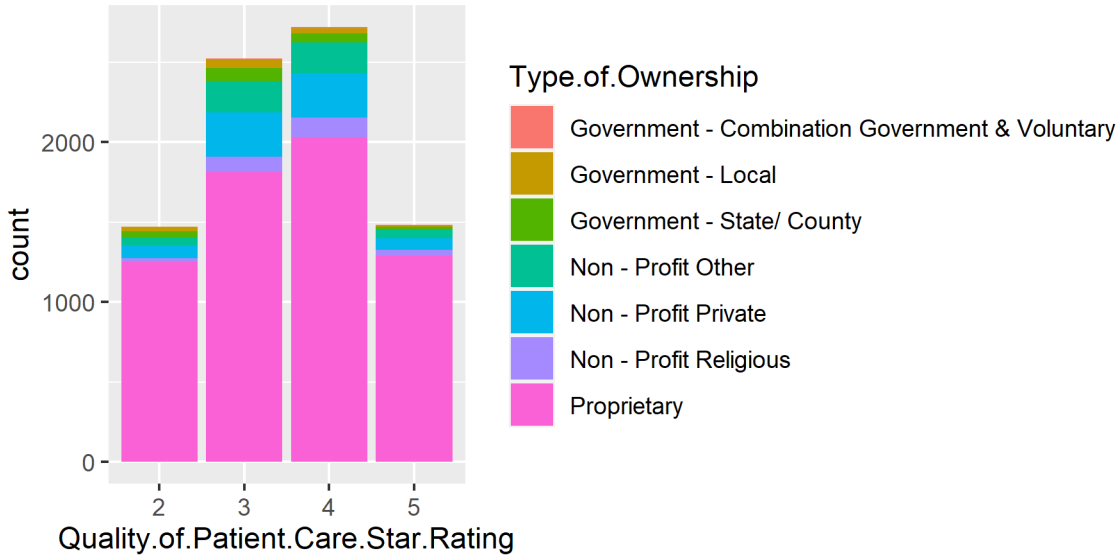
The outcome I am interested in is the `Quality of patient care star rating`. Although this variable is collected as a numeric rating from 1 through 5 in increments of 0.5, I treat it categorically with 9 possible outcomes.

In this report, I am mainly concerned with how those continuous variables are correlated with the quality rating. My effort is put on building up a model for prediction. For convenience, the variables name in this report will be changed as in the Appendix A.

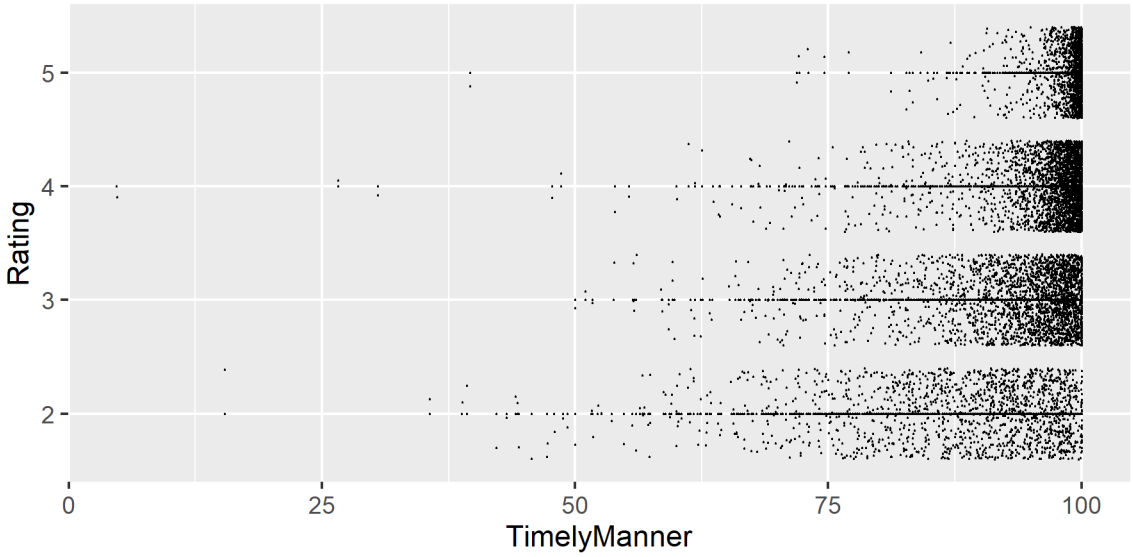
Method

Exploratory Data Analysis

In this section, I will apply EDA on the dataset and get a sense of certain characteristics of the data. To simplify the the analysis and accelerate model fitting, I will take ceiling of the `Quality of patient care star rating`.
First, I want to get a sense of the distribution of the quality rating by type of ownership. Since the observations from the first category is much less than the other categories, the rating 1 and 2 are combined as one category and denoted as `2`



The `Proprietary` occupies the largest proportion in all the four outcome categories, and the other types look quite evenly distributed. The EDA plot is not imformative enough, which leads us to look into modelling in the next section. Plot for continuous variables against `Rating` is as follows:

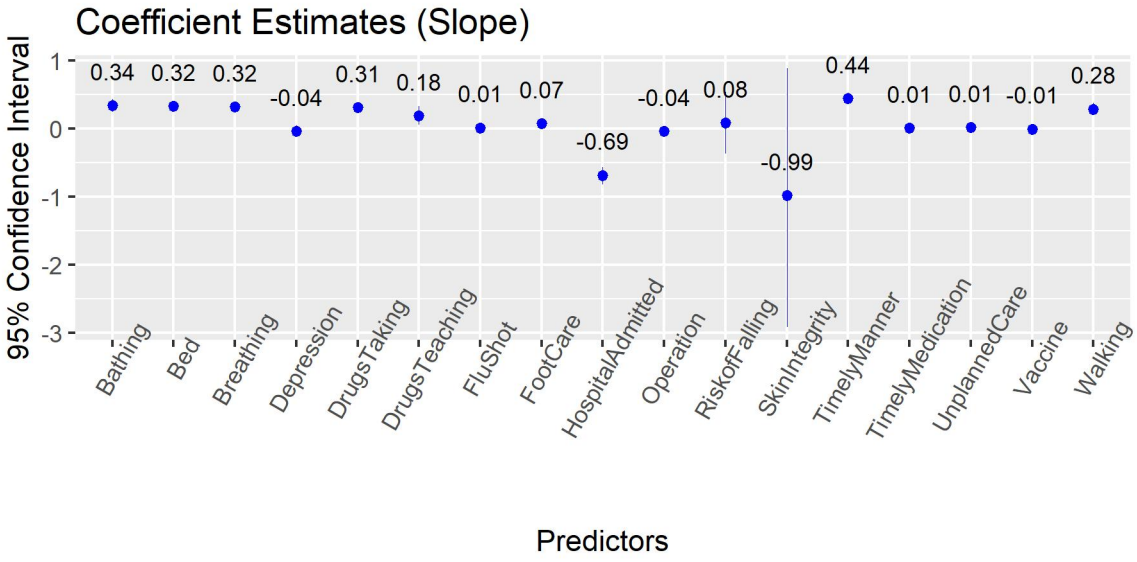


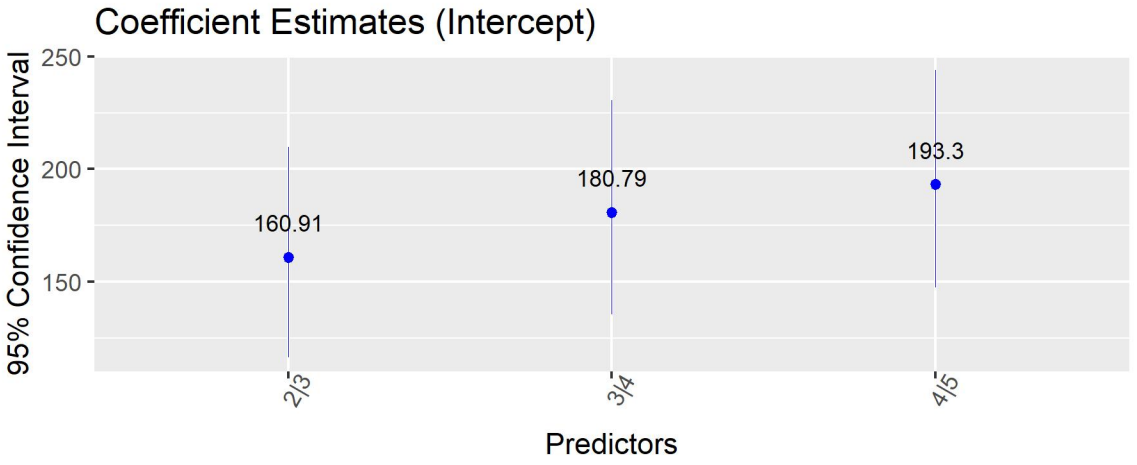
An example of continuous variables versus `Quality of Patient Care Star Rating` are plotted above. The rest will be put in the Appendix B.

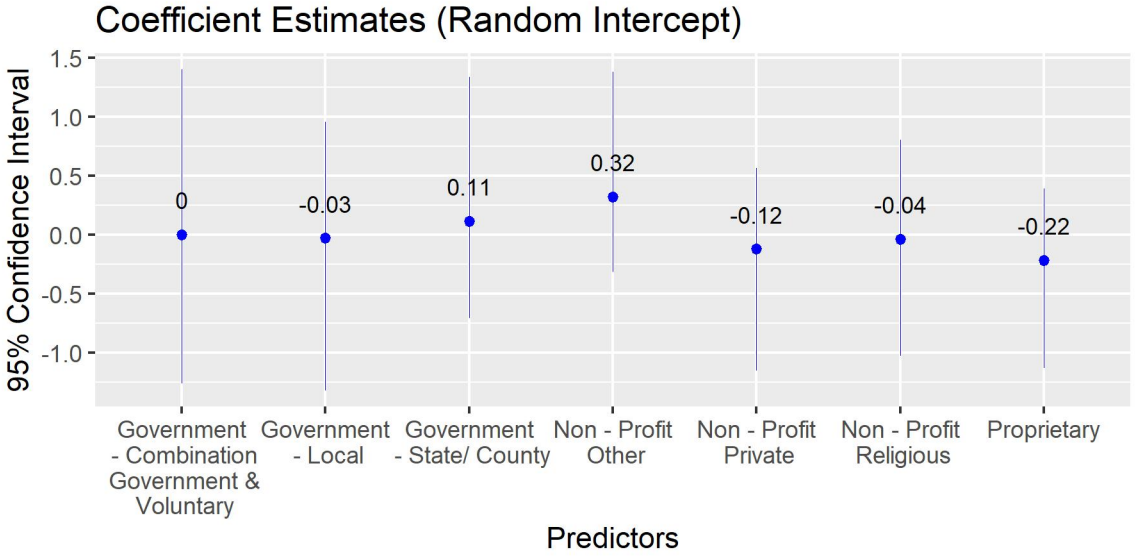
Modelling

Multilevel Ordinal Logistic Regression Model

In this section, I will fit a multilevel ordinal logistic regression model on the dataset. I will use `Type.of.Ownership` as random effects, in order to looking at the ratings between different types of ownerships, figuring out if there is a difference between them. The `RStan` package will be used. The 95% confidence intervals of the coefficient estimates are shown in the following plot:

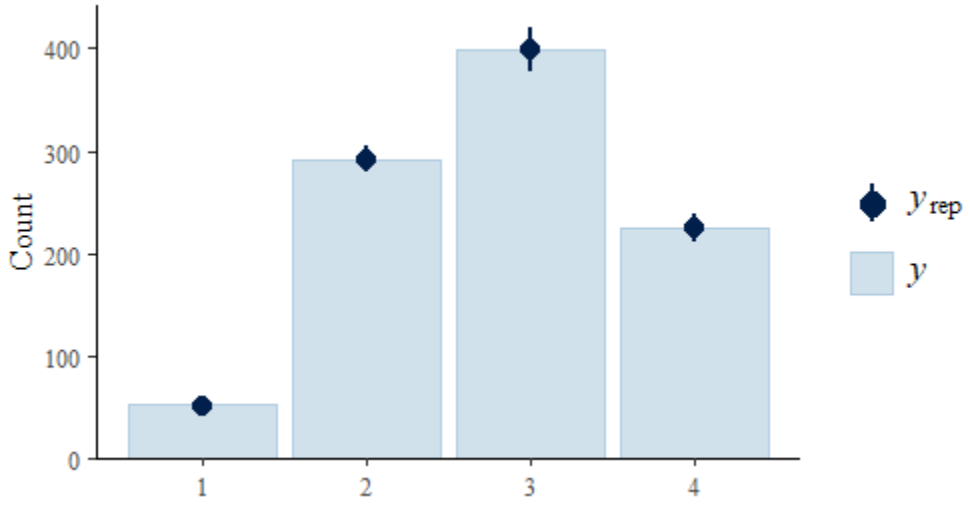

Coefficient Estimates (Slope)
95% Confidence Interval
1
0
-1
-2
-3
0.34
0.32
0.32
-0.04
0.31
0.18
0.01
0.07
-0.69
-0.04
0.08
-0.99
0.44
0.01
0.01
-0.01
0.28
Bathing
Bed
Breathing
Depression
DrugsTaking
DrugsTeaching
FluShot
FootCare
HospitalAdmitted
Operation
RiskofFalling
SkinIntegrity
TimelyManner
TimelyMedication
UnplannedCare
Vaccine
Walking
Predictors


Coefficient Estimates (Intercept)
95% Confidence Interval
250
200
150
160.91
180.79
193.3
2|3
3|4
4|5
Predictors


Coefficient Estimates (Random Intercept)
95% Confidence Interval
1.5
1.0
0.5
0.0
-0.5
-1.0
0
-0.03
0.11
0.32
-0.12
-0.04
-0.22
Government - Combination Government & Voluntary
Government - Local
Government - State/ County
Non - Profit Other
Non - Profit Private
Non - Profit Religious
Proprietary
Predictors

Validation

In this section, I will validate the model by using posterior predict from the model. By building a predict model in Stan, I generate a ppc bar plot below. In this ppc bar plot, the darker blue dots and lines indicate the medians of posterior samples `yrep` with their confidence intervals.



Result

EDA

From the EDA part, there are two points to be stressed on. First, the barplot shows that counts of rating on both ends are less than the middle ones, which accords with common sense. Second, in the scatter point plot, for variables `RiskofFalling`, `DrugsTeaching`, `RiskofFalling`, `Depression`, `FootCare`, the points concentrate on the right side, indicating these services are usually provided with a large percentage. Variables `FluShot` and `Vaccine`, while also show a tendency on a large percentage of service, they also have a larger dispersion towards left. In addition, the dots also have a tendency of concentrating on the upper-right, indicating that there should be some positive correlation betIen the variables and the ratings.
Moreover, `Walking`, `Bed`, `Bathing`, `Breathing`, `Operation`, `DrugsTaking` are showing a more obvious positive correlation. The `HospitalAdmitted` and `SkinIntegrity` show negative correlations, which is consistent with what the variables represent.

Model

(i) The `SkinIntegrity` have a large standard error, indicating that there is a large uncertainty in the estimate.

(ii) Several coefficients, such as `Depression`, `Flushot`,`Vaccine`,`Operation`,`UnplannedCare`,`TimelyMedication` and `RiskofFalling`, not only have estimates very close to 0, but also have 95% confidence interval crossing 0. This result indicates that we can not safely reject that these coefficients should be 0. The result is also consistent with what I found in the EDA, where under these predictors, the `Rating` doesn't show a noticable positive or negative correlation.

(iii) `Bathing`, `Bed`, `Breathing`, `DrugsTaking`, `TimelyManner` and `Walking` show significant estimates, and their estimates are relatively high, indicating these variables are showing stronger positive correlation with the outcome variable. The `HospitalAdmitted`, on the opposite, shows a strong negative correlation with the outcome variable, which is also consistent with what the variable means: How often home health patients had to be admitted to the hospital.

(iv) In the random intercept plot, all the coefficient estimates have confidence intervals crossing 0, indicating there might not be a 95% significant difference between `Types of Ownership`. The `Non - Profit Other` group have a relatively high estimate, indicating that they might not have a higher rating when the predictors are controlled as the same comparing with other groups, but with an insignificant result it is not a safe conclusion to draw.

Validation

In the validation part, I use ppc bar plot to show how well the model fits. From the plot, the medians of predictive values `yrep` are basically the same as the counting of observative values `y`, indicating the model fits well on the datset

Discussion

The results drawn above shows that, for some variables, they might not have siginificant impacts on the `Rating`, such as `Depression` and all those insignificant variables in 5% level shown in the results. For those significant ones, there are two groups of them. First, for those having low estimates, they are showing a positive correlation with `Rating`, but they contribute small to the increase on the `Rating`. Meanwhile, for those having high estimates, the model predicts a larger increase on `Rating` when they are increasing. Therefore, the team should expect a higher `Rating` from the patient, if they have done well in the following aspects:
Help patients get better at bathing (`Bathing`);
Help patients get in and out of bed (`Bed`);
Help patients get better at breathing (`Breathing`);
Help patients get better at taking their drugs correctly by mouth (`DrugsTaking`);
Begin patient's care in a timely manner (`TimelyManner`);
Help pateints get better at walking or moving around (`Walking`).
Also, if the home health team successfully reduce the number of having the patients admitting to the hospital (`HospitalAdmitted`), they should expect a decent increase on the `Rating` as well. Furthermore, I want to stress that the analysis above does not guarantee any causal relationship. For example, if a team tries to improve their rating by purposely avoid admitting patients into the hospital even when they should, they will not be able to see any improvement in the rating.

References

[1] Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge university press.
[2] Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2. http://mc-stan.org/.

Appendix A

`Rating`: Quality of patient care star rating, a numeric rating from 1 through 5, in increments of 0.5. Factored in this report.

`TimelyManner` : How often the home health team began their patients' care in a timely manner.

`DrugsTeaching`: How often the home health team taught patients (or their family caregivers) about their drugs

`RiskofFalling`: How often the home health team checked patients' risk of falling

`Depression`: How often the home health team checked patients for depression

`FluShot`: How often the home health team determined whether patients received a flu shot for the current flu season

`Vaccine`: How often the home health team made sure that their patients have received a pneumococcal vaccine (pneumonia shot)

`FootCare`: With diabetes, how often the home health team got doctor's orders, gave foot care, and taught patients about foot care

`Walking`: How often patients got better at walking or moving around

`Bed`: How often patients got better at getting in and out of bed

`Bathing`: How often patients got better at bathing

`Breathing`: How often patients' breathing improved

`Operation`: How often patients' wounds improved or healed after an operation

`DrugsTaking`: How often patients got better at taking their drugs correctly by mouth

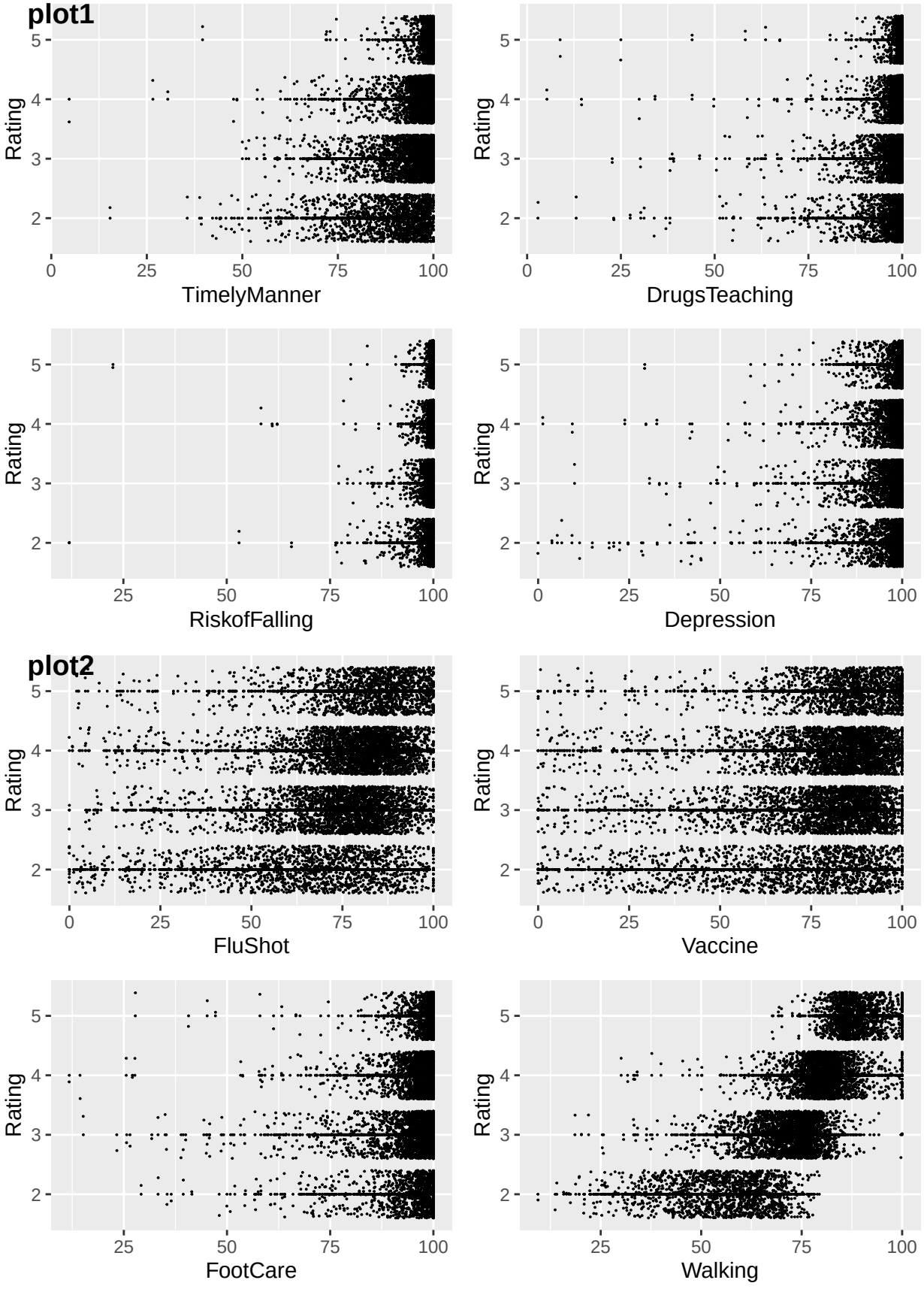
`HospitalAdmitted`: How often home health patients had to be admitted to the hospital

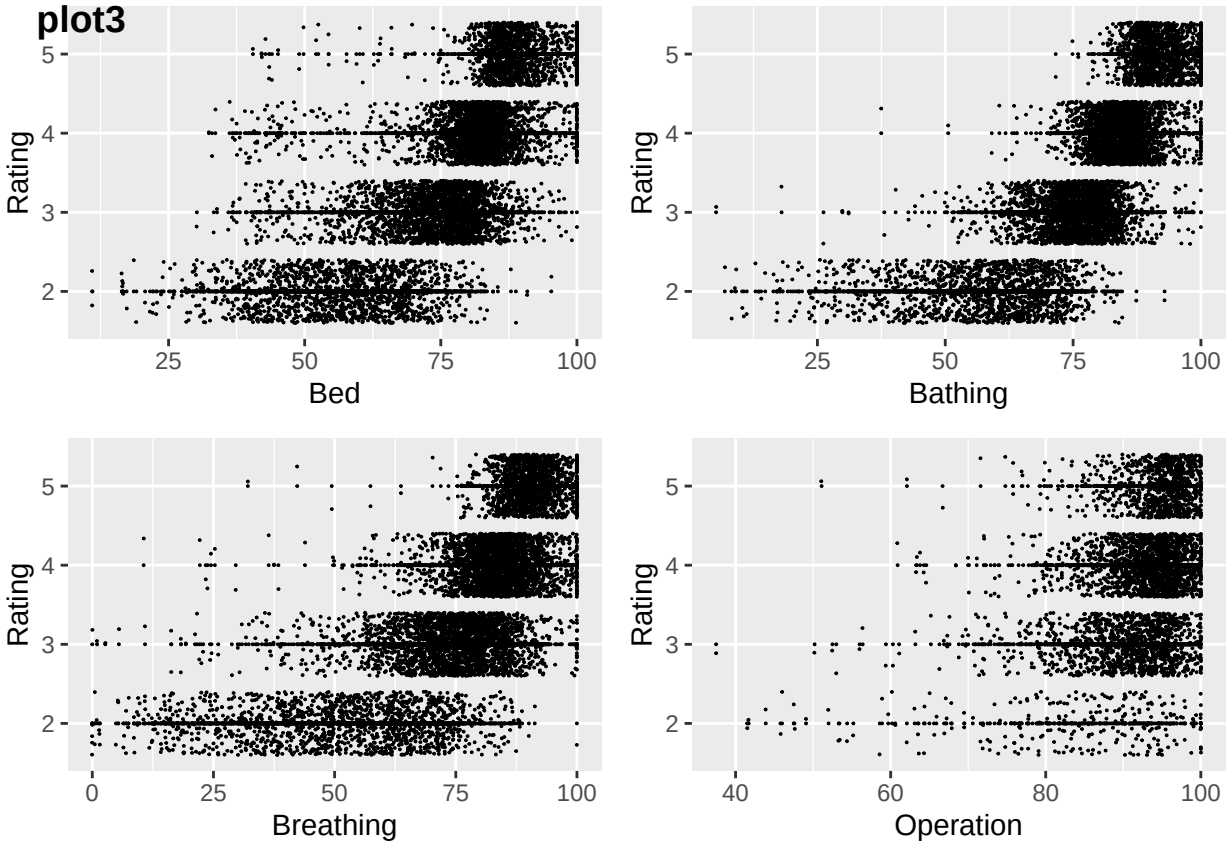
`UnplannedCare`: How often patients receiving home health care needed urgent, unplanned care in the ER without being admitted

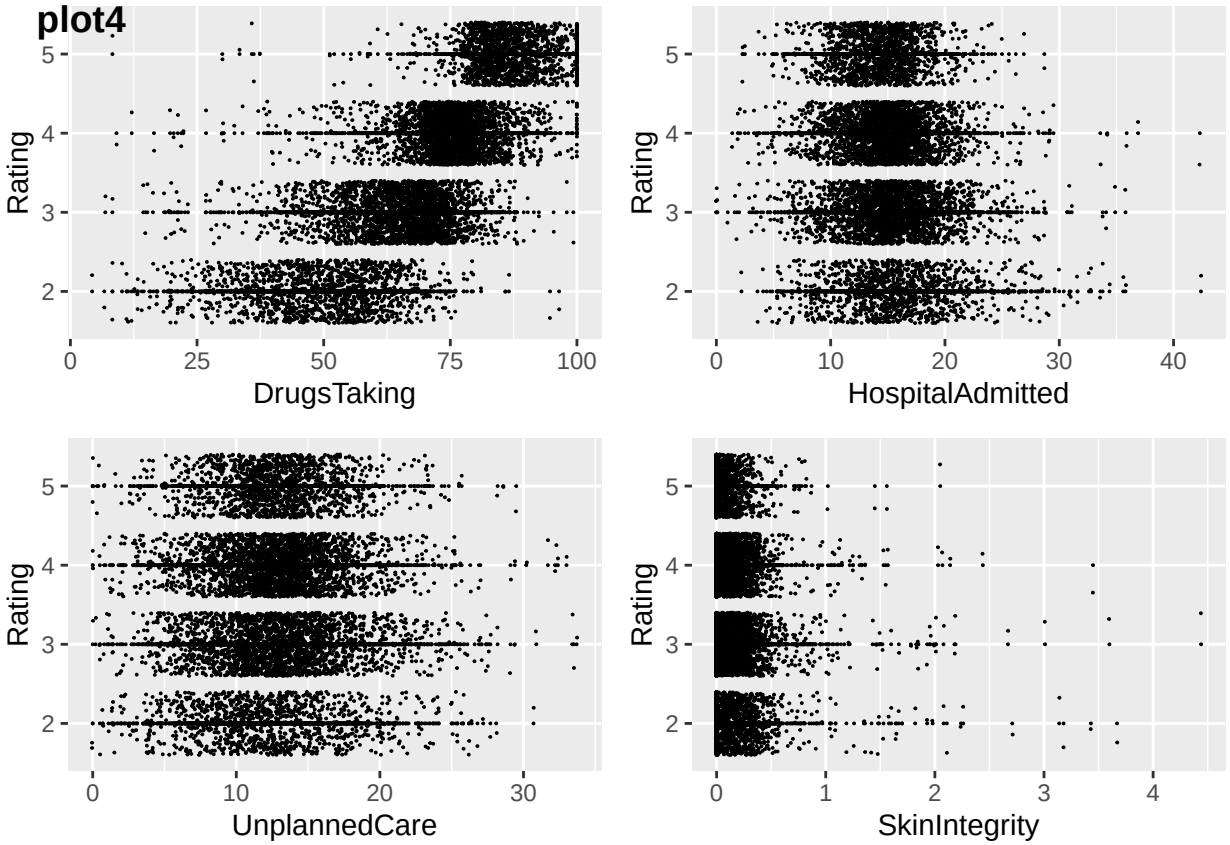
`SkinIntegrity`: Changes in skin integrity post-acute care: pressure ulcer/injury

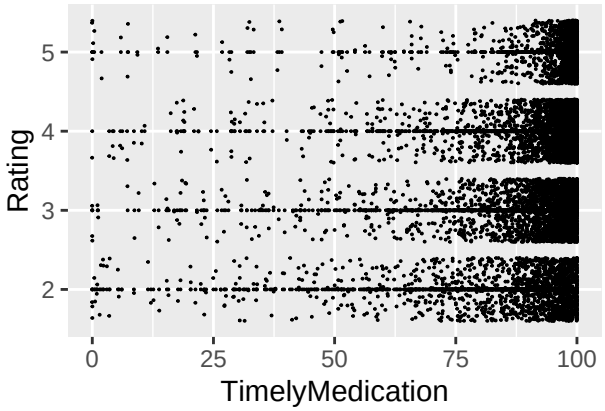
`TimelyMedication`: How often physician-recommended actions to address medication issues Ire completely timely

Appendix B: EDA Plots




plot3
Rating
5
4
3
2
25
50
75
100
Bed
Bathing
0
Breathing
40
60
80
Operation


plot4
Rating
DrugsTaking
HospitalAdmitted
UnplannedCare
SkinIntegrity


Rating
5
4
3
2
0
25
50
75
100
TimelyMedication