

STAT 6021 Project2 (Group 2) Proposal: Characteristics of nursing homes in New Mexico.

Gregory Madden, Christina Kuang, Chi Do, Trey Hamilton

Part 2: Proposal

Names of the R package and the R dataframe:

R Package: “*data(”Nursing“)*”

R dataframe: “*Nursing*”

Description of Dataset (provided by Stat2Data Package)

Characteristics of nursing homes in New Mexico.

Format: A dataset with 52 observations on the following 7 variables.

Beds Number of beds in the nursing home InPatientDays Annual medical in-patient days (in hundreds)
AllPatientDays Annual total patient days (in hundreds) PatientRevenue Annual patient care revenue (in
hundreds of dollars) NurseSalaries Annual nursing salaries (in hundreds of dollars) FacilitiesExpend Annual
facilities expenditure (in hundreds of dollars) Rural 1=rural or 0=non-rural

Details: The data were collected by the Department of Health and Social Services of the State of New Mexico and cover 52 of the 60 licensed nursing facilities in New Mexico in 1988.

References: Howard L. Smith, Niell F. Piland, and Nancy Fisher, “A Comparison of Financial Performance, Organizational Characteristics, and Management Strategy Among Rural and Urban Nursing Facilities,” Journal of Rural Health, Winter 1992, pp 27-40.

A link to the data set:

Data can be found at: <http://lib.stat.cmu.edu/DASL/Datafiles/Nursingdat.html>

Saving a file containing the data, as a .csv file.

```
# For some reason, Nursing dataset does not call correctly when knitting
# markdown so commented here uncomment and run below code when running in new
# environment

# Data <- Nursing write.csv(Data, 'Nursing.csv')

Data <- read.csv("Nursing.csv")
```

Showing the first few entries of the dataset:

```
head(Data)
```

```
##      X Beds InPatientDays AllPatientDays PatientRevenue NurseSalaries
## 1 1   244          128          385          23521          5230
## 2 2    59          155          203          9160          2459
```

##	3	3	120	281	392	21900	6304
##	4	4	120	291	419	22354	6590
##	5	5	120	238	363	17421	5362
##	6	6	65	180	234	10531	3622
##			FacilitiesExpend	Rural			
##	1		5334	0			
##	2		493	1			
##	3		6115	0			
##	4		6346	0			
##	5		6225	0			
##	6		449	1			

- Project objectives/goals. What questions is the group trying to answer, as well as potential practical implications (the more interesting and/or practical, the better) of the results?

Q1. What characteristics of nursing homes in New Mexico dictate annual nurse salaries at those institutions?

Practical implications of a linear model for predicting nurse salaries could be used by policymakers to rationally distribute subsidy funds to institutions that are expected to have the lowest salaries.

Q2. What factors are associated with rural vs. non-rural nursing homes?

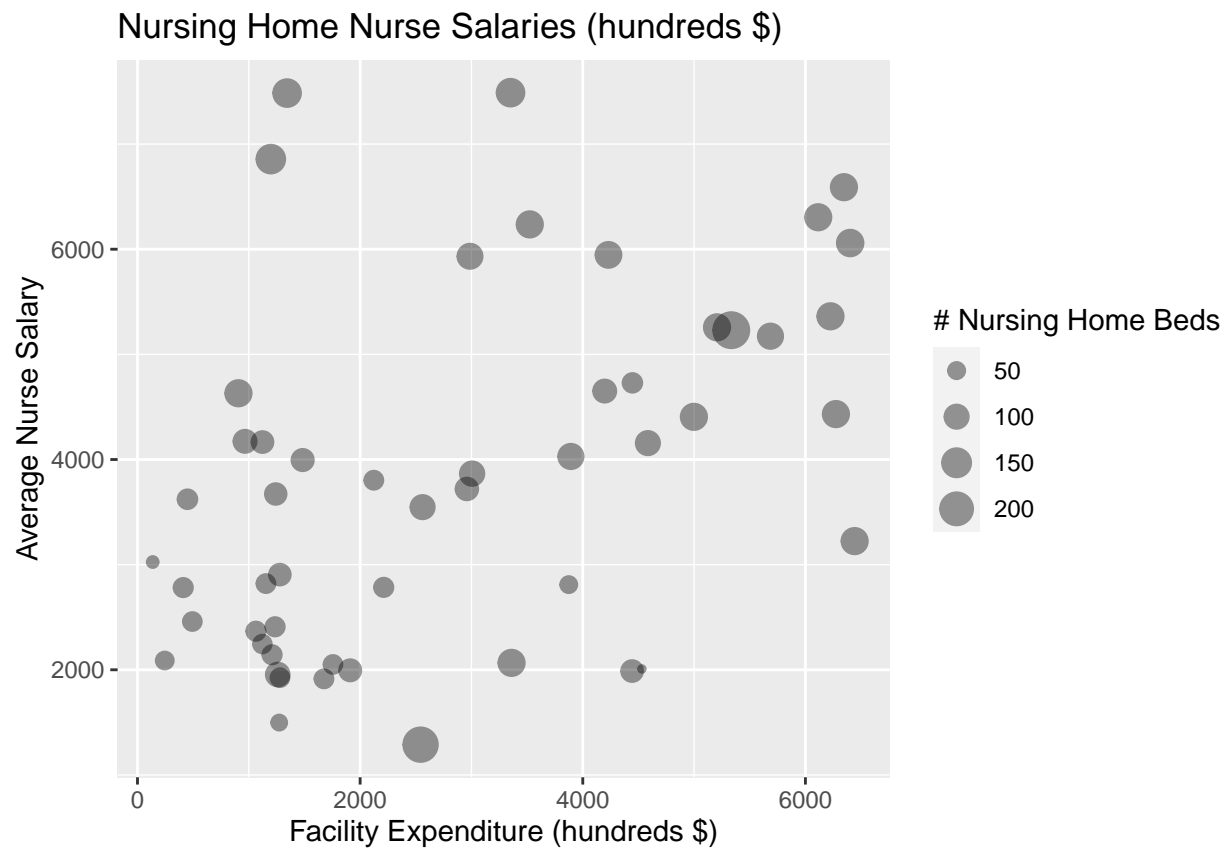
Rural patients suffer from a lack of locally available nursing home beds. Understanding the relationships of these characteristics and how they define rural vs. non-rural nursing homes is helpful to know how to make rural nursing homes financially viable.

- Some data visualizations and commentary related to the project objectives/goals.

Nursing home salaries by Annual facilities expenditure (in hundreds of dollars) and Nursing home size (by number of beds):

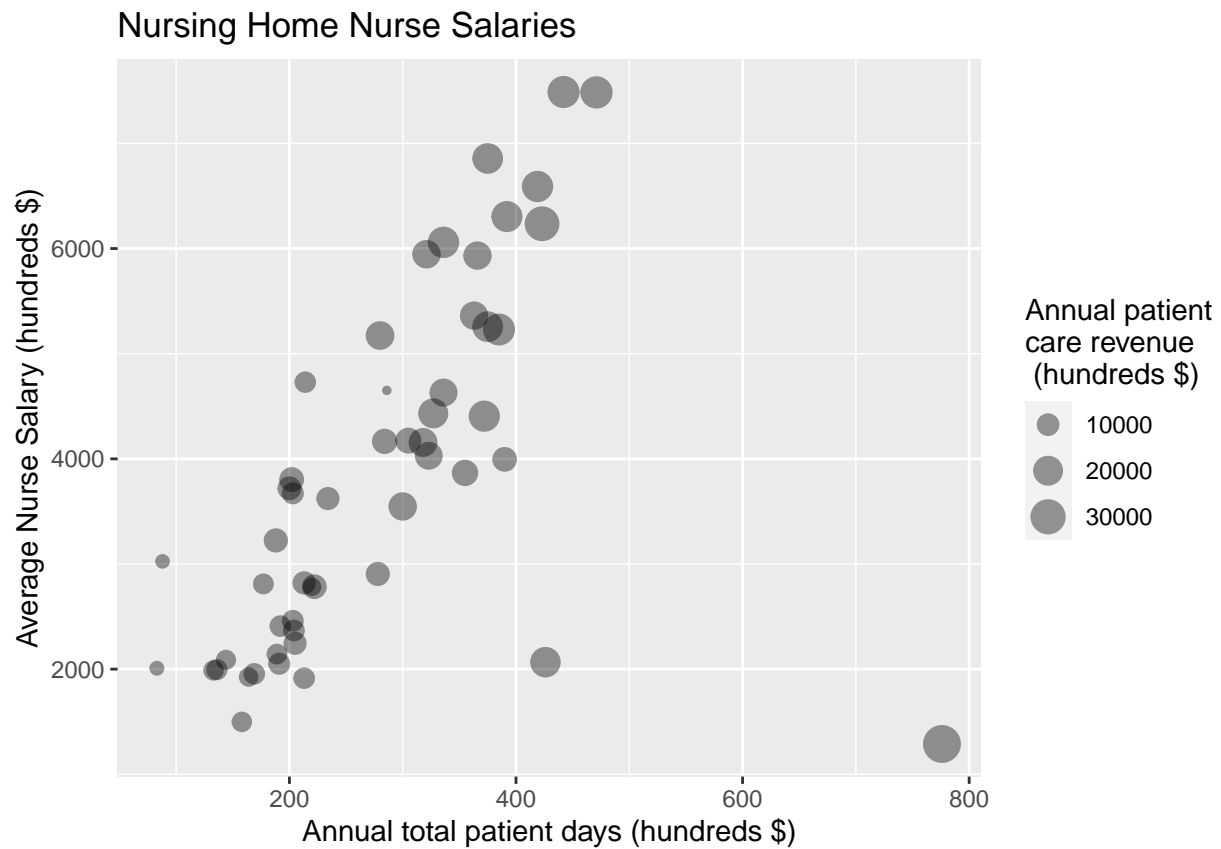
Data %>%

```
ggplot(aes(x = FacilitiesExpend, y = NurseSalaries, size = Beds)) + geom_point(alpha = 0.4) +
  labs(x = "Facility Expenditure (hundreds $)", y = "Average Nurse Salary", title = "Nursing Home Nur") +
  guides(size = guide_legend(title = "# Nursing Home Beds"))
```



Nursing home salaries by census in Annual total patient days (in hundreds) and Nursing home size (by number of beds):

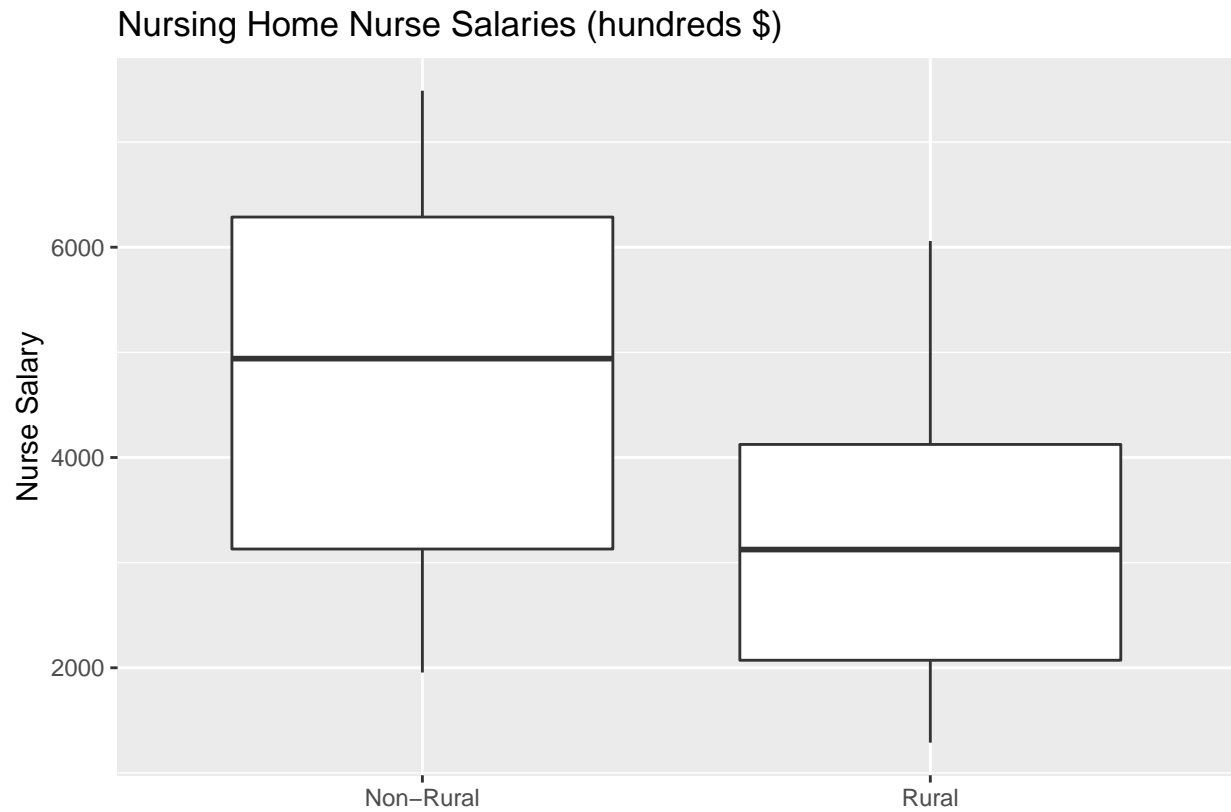
```
Data %>%
  ggplot(aes(x = AllPatientDays, y = NurseSalaries, size = PatientRevenue)) + geom_point(alpha = 0.4)
  labs(x = "Annual total patient days (hundreds $)", y = "Average Nurse Salary (hundreds $)",
        title = "Nursing Home Nurse Salaries") + guides(size = guide_legend(title = "Annual patient \ncensus"))
```



Boxplot demonstrating differences in Institutional nurse's salaries in New Mexico for Rural Areas compared with Non-Rural:

```
Data$Rural <- factor(Data$Rural)
levels(Data$Rural) <- c("Non-Rural", "Rural")

Data %>%
  ggplot(aes(x = Rural, y = NurseSalaries)) + geom_boxplot() + labs(x = "", y = "Nurse Salary",
    title = "Nursing Home Nurse Salaries (hundreds $)")
```



Objectives:

Objective 1: Fit a multiple linear regression model with annual nurse salary for individual nursing homes using the available financial characteristics for each institution. The goal is to develop a model using these available data to reliably predict annual nursing salaries among the larger group of nursing homes across the state.

Objective 2: Fit a logistic regression model with rural vs. non-rural status as the response using the available financial characteristics.