Report of Oral and Facial Study: Visualization and Interpretation

Tianjian Xie, Amstrong Wang, Haochen Pan, Nuo Chen, Haochen Zhu

2022-10-28

Introduction

Our client is David Cartier from Boston University Henry M. Goldman School of Dental Medicine. The client's goal is to analyze the different types of facial fractures from Maine Portland medical center and compare injuries between rural and urban populations.

The client asked us to compare two different injury assessment scores between rural and urban populations. These scores are the injury severity score (ISS) and Glasgow coma scale (GCS). The client also asked us to assess the relationship between injury mechanisms and these two scores, as well as explore factors affecting the frequency of hospital admissions.

Data Description

The data consists of de-identified medical records for 318 patients from Maine Medical Center. Patients are classified as coming from rural areas or urban areas based on their Zip code and 2010 census data.

The Glasgow Coma Scale (GCS) is used to objectively describe the extent of impaired consciousness in all types of acute medical and trauma patients. The scale assesses patients according to three aspects of responsiveness: eye-opening, motor, and verbal responses. This scale is from 1 to 15 which 13-15 being Mild, 9-12 being Moderate, and 3-8 being Severe.

All injuries are assigned from an internationally recognized dictionary that describes over 2000 injuries. Multiple injuries are scored by adding together the squares of the three highest AIS scores. The ISS is the Injury severity score, which can range from 1 to 75, but we only have from 1 to 50.

Exploratory Data Analysis

Data Visualization: 1. ISS vs GCS

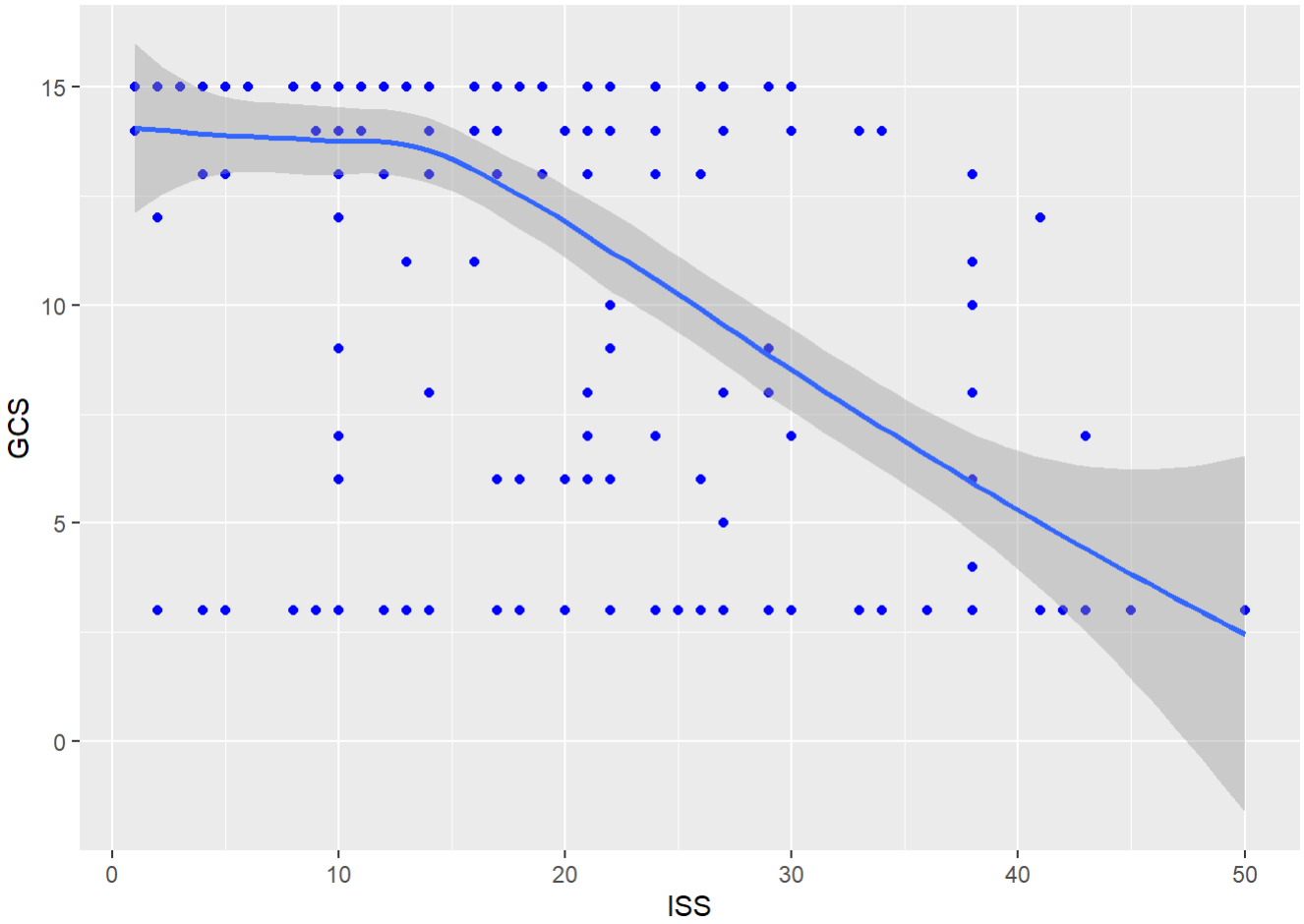

Figure 1: ISS Against GCS

Figure 1 shows that the patients' GCS are concentrated on top and bottom, and the smooth line is not linear. It can not prove our client's assumption that there is an inversely proportional correlation between ISS and GCS very well.

2. Total.Hospital.Days vs ISS

```
## NULL
```

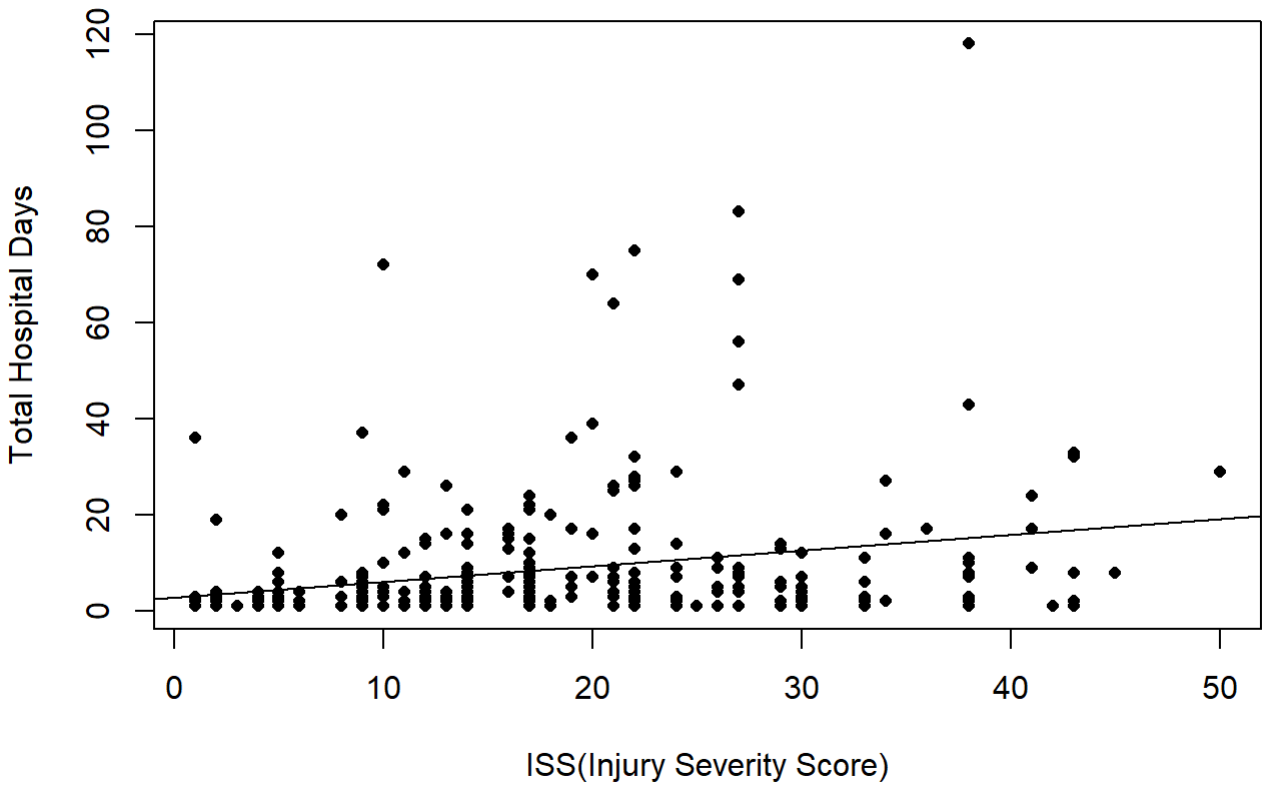

Figure 2: Total.Hospital.Days vs ISS

Figure 2 shows the correlation between the total days patients stay in hospital and their ISS. The fitting line shows there is a directional proportional correlation between patients' ISS and total days they spent in hospital.

3. Urban or Rural for ISS

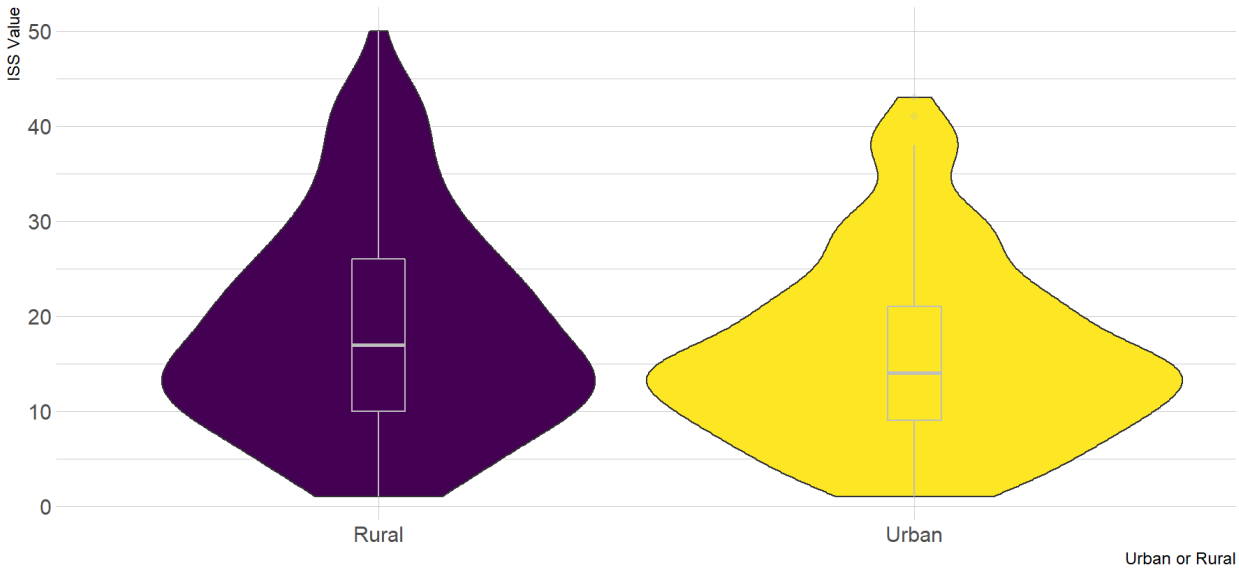

Figure 3: Violin Plot of Urban or Rural for ISS

Figure 3 shows for ISS, urban and rural patients are likely to have similar ISS, and the means for both region are around 15. Compare to the rural patients, the urban patients' ISS are more concentrate to the mean since the violin are wider.

4. Urban or Rural for GCS

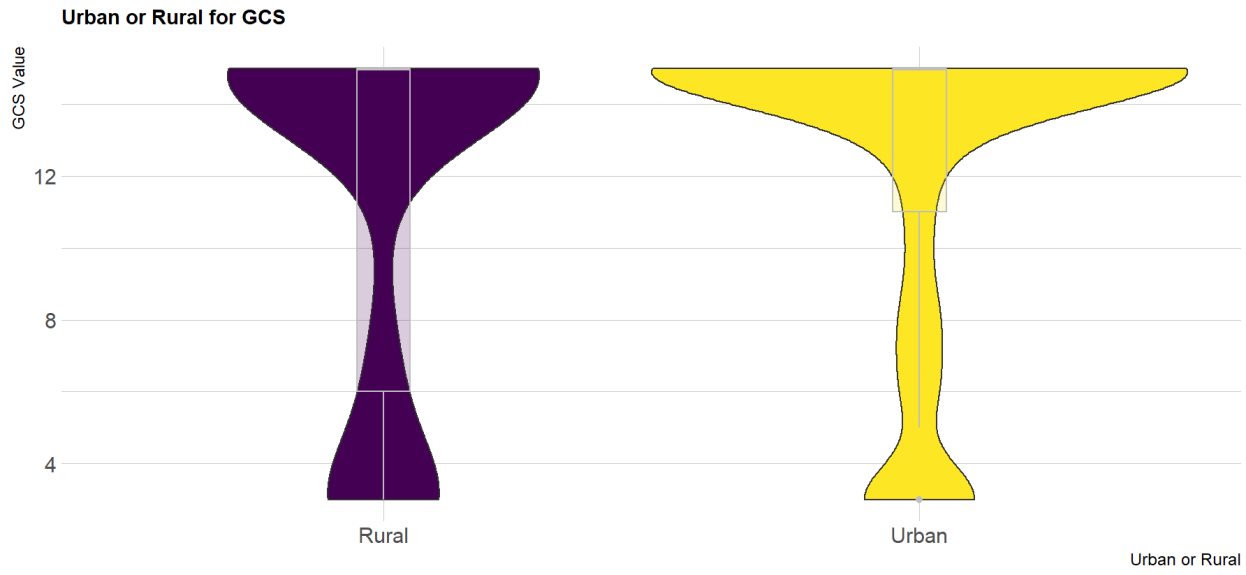

Figure 4: Violin Plot of Urban or Rural for GCS

From figure 4, we can see that most of the patients' GCS are concentrated near 15, which is the top of the graph, and corresponding to less-severe injury. The concentration is somewhat larger for Urban compared to Rural. The explanation can be that people close to Maine Medical Center might go to Maine Medical Center even if they have less severe injuries, while people further away might not bother traveling that distance if they have less severe injuries.

Methods

1. Fitting model for ISS and GCS

```
##
## Call:
## lm(formula = GCS ~ ISS, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.667  -1.720   1.049   2.276   6.199
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.15777    0.44133   36.61   <2e-16 ***
## ISS         -0.24524    0.02185  -11.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.996 on 314 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.2863, Adjusted R-squared:  0.284
## F-statistic: 125.9 on 1 and 314 DF,  p-value: < 2.2e-16
```

The summary of the model shows the estimate GCS of a patient with 0 ISS is 16.15777, and with every unit increase in ISS, the estimate GCS decrease by 0.24524. So with higher ISS, the patient is expected to have lower GCS. This shows a negative relationship between ISS and GCS.

2. Fitting model for ISS and total days in hospital

```
##
## Call:
## lm(formula = Total.Hospital.Days ~ ISS, data = cleaned_data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.857  -5.780  -3.504   0.202 102.768
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.87866    1.49082   1.931   0.0544 .
## ISS          0.32508    0.07382   4.404 1.46e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.5 on 314 degrees of freedom
## Multiple R-squared:  0.05817,    Adjusted R-squared:  0.05517
## F-statistic: 19.39 on 1 and 314 DF,  p-value: 1.461e-05
```

The summary of the model shows the estimate of the days a patient with 0 ISS is 2.87866, and with every unit increase in ISS, the estimate day increase by 0.32508. So with higher ISS, the patient is expected to stay longer in hospital.

3. Fitting model for ISS and Urban or Rural and make T test

```
##
## Call:
## lm(formula = ISS1 ~ UR)
##
## Coefficients:
## (Intercept)      URUrban
##      16.230       -2.098
```

```
##
##  Welch Two Sample t-test
##
## data:  UrbISS and RurISS
## t = -2.4233, df = 299.39, p-value = 0.01598
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.0657178 -0.5252778
## sample estimates:
## mean of x mean of y
##  15.99398  18.78947
```

The result of the t test of ISS based on the Urban or Rural.

4. Fitting model for GCS and Urban or Rural and make KS test

```
##
## Call:
## lm(formula = GCS1 ~ UR)
##
## Coefficients:
## (Intercept)      URUrban
##      9.4636       0.8273
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  UrbGCS and RurGCS
## D = 0.10351, p-value = 0.3669
## alternative hypothesis: two-sided
```

Based on the small p-values, we decided that a KS test was more appropriate than a t-test, because a KS test makes fewer assumptions about the underlying distribution.

Since the p-value is larger than 0.05, it's failed to reject the hypothesis. Then, the KS test said that we don't have enough evidence to say whether there was any marginal difference between Rural and Urban populations.

Discussion

Our analysis of patients come from US 2010 census data shows that there is a negative and interactive relationship between ISS value and GCS value, and is related to urban/rural areas. ISS and GCS are normally distributed at a level at 0.05. In our analysis, urban or rural patients may have similar ISS and be more clustered

towards the mean. However, the relationship between GCS and urban/rural cannot be directly explained through a linear relationship since we have evidence that ISS and GCS are not normally distributed. In the comparison of other variables, we also believe that factors such as hospitalization time and cause of injury have no way to directly affect the GCS value in different urban/rural situations.

Conclusion

Our linear model suggests that there isn’t evidence that there is a relationship between GCS and whether a patient is from a rural or Urban origin. In our analysis, urban or rural patients may have similar ISS and be more clustered towards the mean. However, the linear model suggests that the is not much evidence of a relationship between GCS and whether a patient’s origin is urban or rural. Although the EDA suggests that there may be some relationship between these scores and whether a patient is from a rural or urban origin, our models suggest that there is not enough evidence to conclude that there is such a relationship.