EDA Project 2

Lance Price

November 28, 2016

I will mainly look at two variables of interest.

One is the measureID and another is the score

For the measure ID, I was only interested in two values:

Acute Myocardial Infarction (AMI) 30-Day Mortality Rate

Acute Myocardial Infarction (AMI) 30-Day Readmission Rate

The score attribute give the values of these rates

Each value was measured for each hospital in the dataset (though not all hospitals have data for these scores).

There some important things to note about these "rates".

All data was subject to the following restrictions:

Only included if:

Enrolled in Medicare fee-for-service (FFS) Part A and Part B for the 12 months prior to the date of the index admission, and enrolled in Part A during the index admission; Aged 65 or over; Not transferred from another acute care facility.

Excluded if:

With inconsistent or unknown vital status or other unreliable demographic (age and gender) data; Enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission; Discharged against medical advice (AMA).

Thus any analysis that is made past this point are specifically relevent to people 65 and older in the medicare program that were not already judged to be close enough to death that they were in a hospice program.

Risk-Adjusted Mortality Rate

The measures assess mortality within a 30-day period from the date of the index admission. The measures use a 30-day time frame because older adult patients are more vulnerable to adverse health outcomes occurring during this time. Death within 30 days of admission can be influenced by hospital care and the early transition to the non-acute care setting. The 30-day time frame is a clinically meaningful period for hospitals to collaborate with their communities in an effort to reduce mortality.

The Mortality rate is calculated as the ratio of the number of "predicted" deaths to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate.

Both the predicted and expected deaths attempt to account for patient-specific criteria that might skew the results and take away from the actual difference between hospitals.

log transformed

heirachical logistic regression models

Extra

The rates use logistic regresssion models. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals. At the patient level, it models the log-odds of mortality within 30 days of the index admission using age, sex (in the AMI, HF, pneumonia, and stroke measures), selected clinical covariates, and a hospital-specific effect. At the hospital level, the approach models the hospital-specific effects as arising from a normal distribution.

The hospital effect represents the underlying risk of mortality at the hospital, after accounting for patient risk. The hospital-specific effects are given a distribution to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital effects should be identical across all hospitals.

The RSMR is calculated as the ratio of the number of "predicted" deaths to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate. For each hospital, the numerator of the ratio is the number of deaths within 30 days predicted based on the hospital's performance with its observed case mix, and the denominator is the number of deaths expected based on the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows a particular hospital's performance, given its case mix, to be compared to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected mortality rates or better quality, while a higher ratio indicates higher-than-expected mortality rates or worse quality. The results are log transformed and summed over all patients attributed to a hospital to get a predicted value. We use the following strategy to calculate the hospital-specific RSMRs, which we calculate as the ratio of a hospital's "predicted" mortality to "expected" mortality multiplied by the national observed mortality rate. The expected mortality for each hospital is estimated using its patient mix and the average hospital-specific effect (that is, the average effect among all hospitals in the sample). The predicted mortality for each hospital is estimated given the same patient mix but an estimated hospital-specific effect. Operationally, the expected number of deaths for each hospital is obtained by summing the expected probabilities of mortality for all patients in the hospital.

Risk-Adjusted Readmission Rate

The measures are designed to capture unplanned readmissions that arise from acute clinical events requiring urgent rehospitalization within 30 days of discharge. Only an unplanned inpatient admission to a short-term acute care hospital can qualify as a readmission. Planned readmissions, which are generally not a signal of quality of care, are not considered readmissions in the measure outcome.

The measures assess unplanned readmissions within a 30-day period from the date of discharge from an index admission. The measures use a 30-day time frame because older adult patients are more vulnerable to adverse health outcomes during this time. Readmission occurring within 30 days of discharge can be influenced by hospital care and the early transition to the non-acute care setting. The 30-day time frame is a clinically meaningful period for hospitals to collaborate with their communities in an effort to reduce readmissions.

Multiple readmissions are only considered as one readmission.

The rates are caluculated in a very similar way to the mortlity rates. The readmission rate is calculated as the ratio of the number of "predicted" readmissions to the number of "expected" readmissions at a given hospital, multiplied by the national observed readmission rate.

measure id was 61.8 % complete

score was also 61.8 % complete

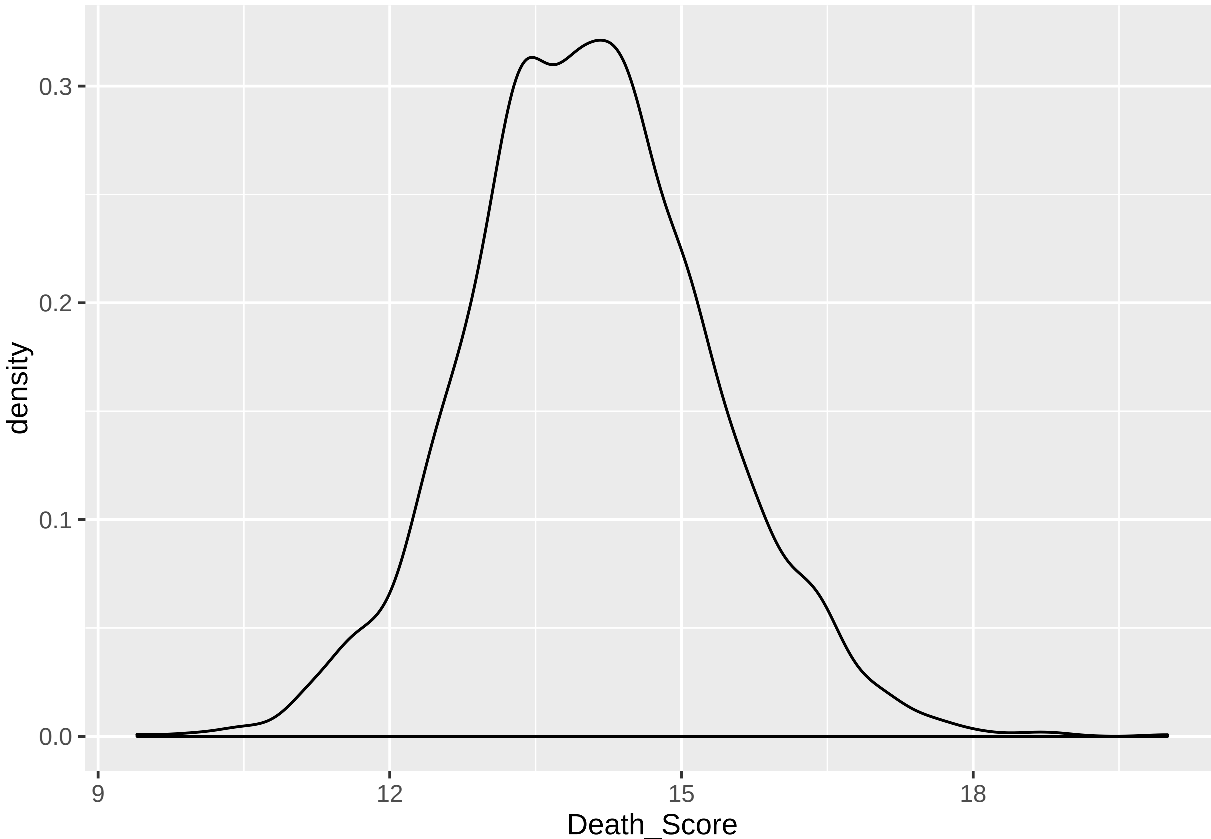
However when limiting the measure ID to mortality rate and readmission rate, the completeness percentages go down to 49.6 % and 44.9 % respectively.

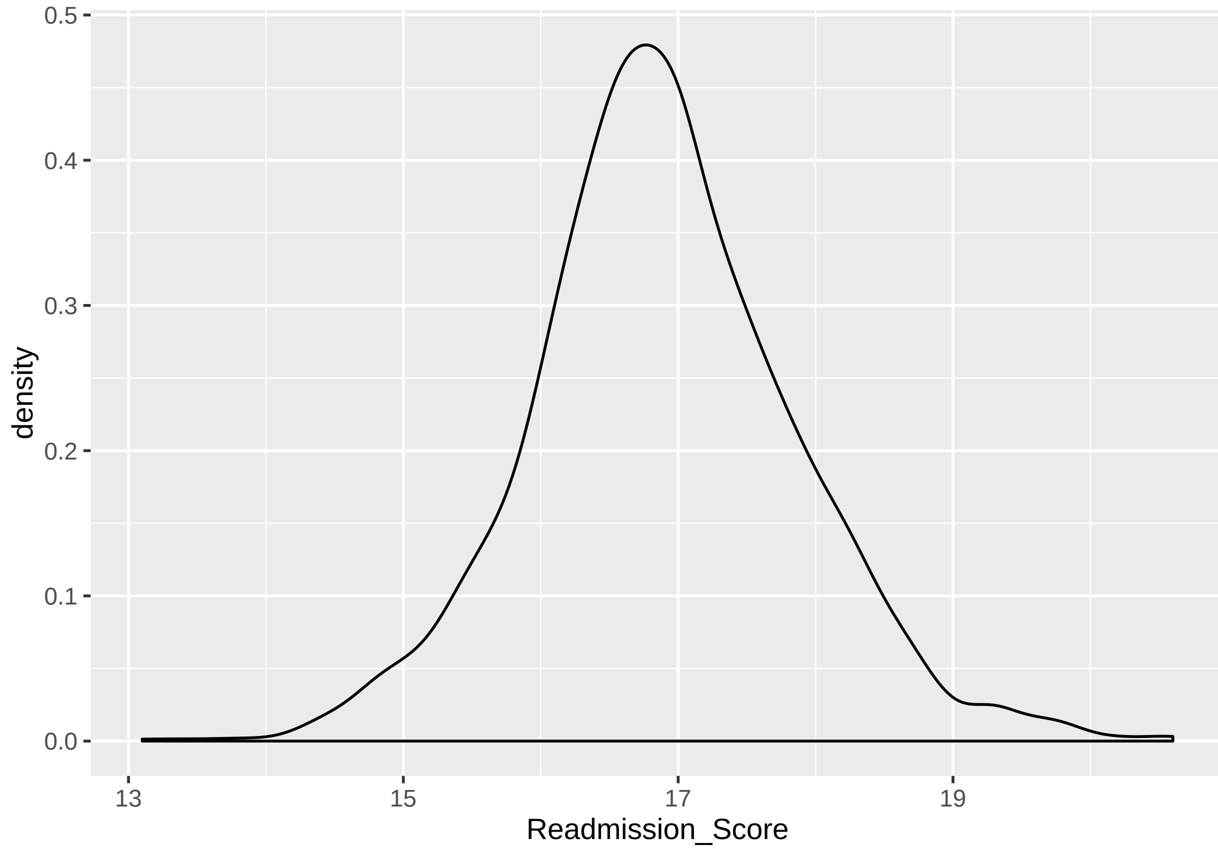
This is somewhat disheartening, but I will move on with the analysis nonetheless because I felt like these were the most informative variables in this particular dataset that could help to answer the prompt that was given.

```
## [1] 61.83657
```

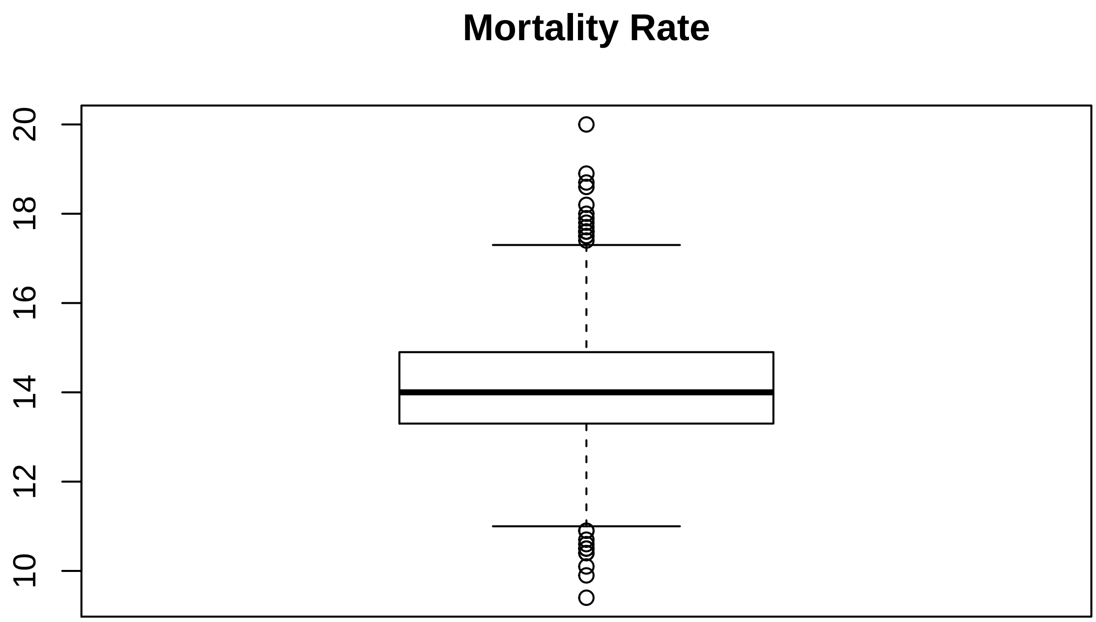
```
## [1] 49.56413
```

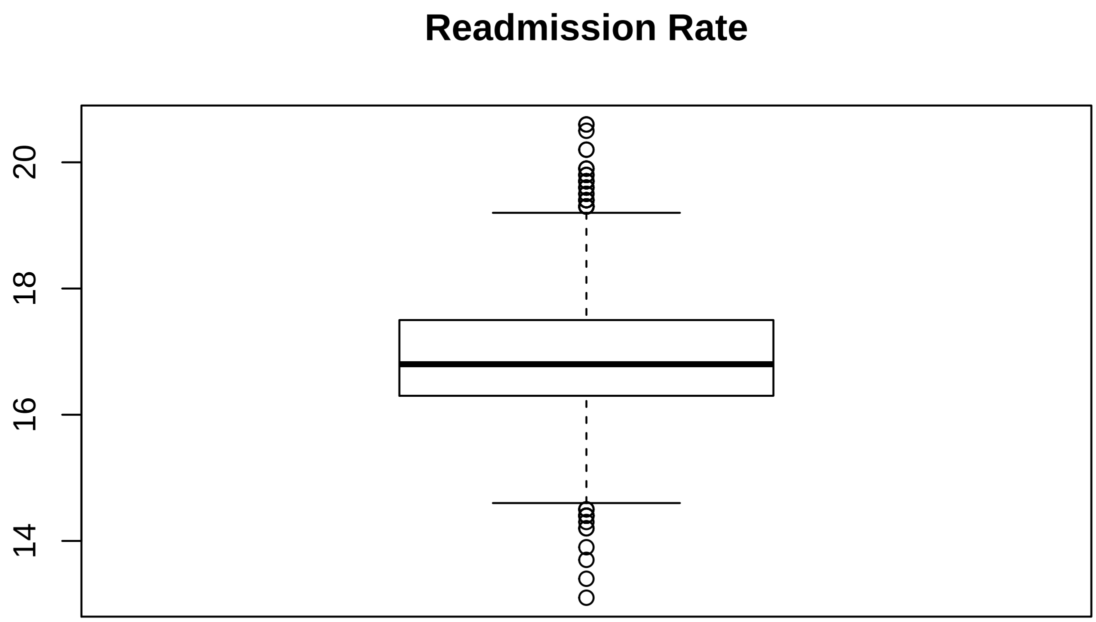
```
## [1] 44.89415
```

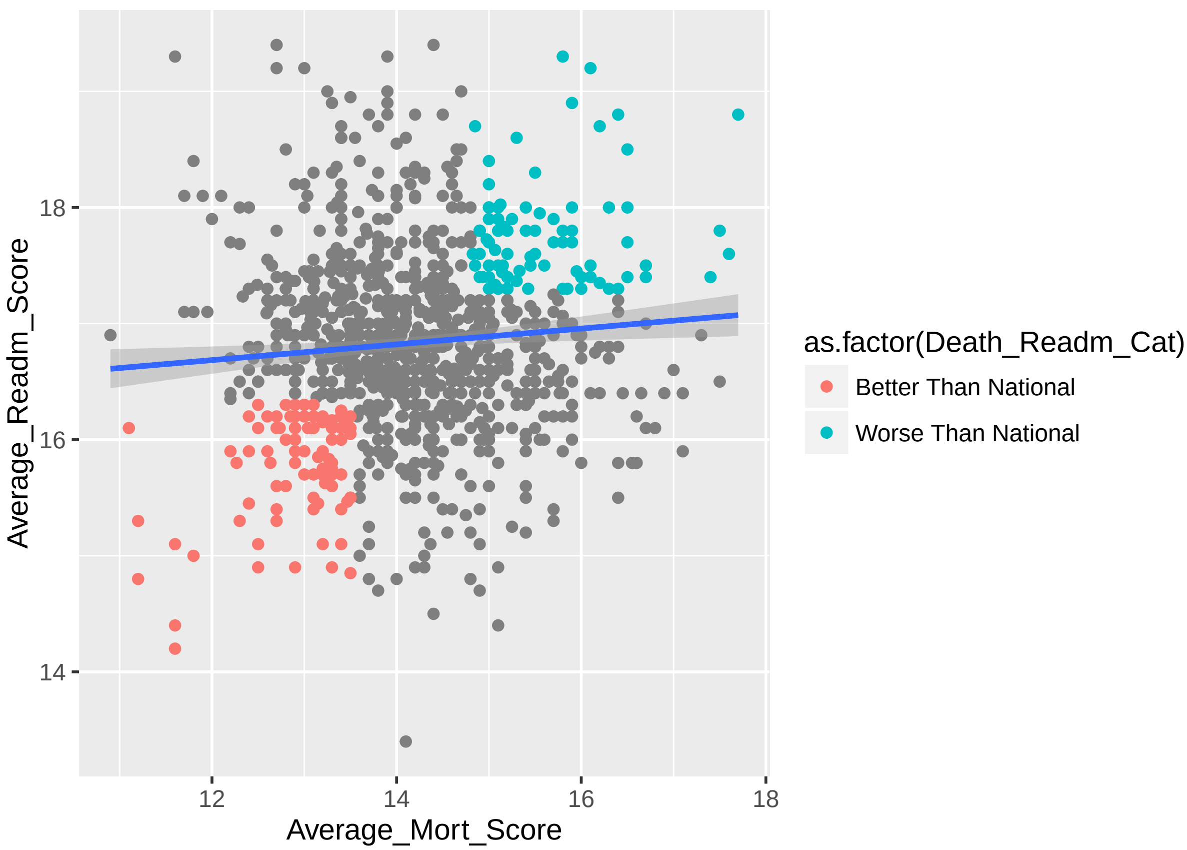

density
0.3
0.2
0.1
0.0
9
12
15
18
Death_Score



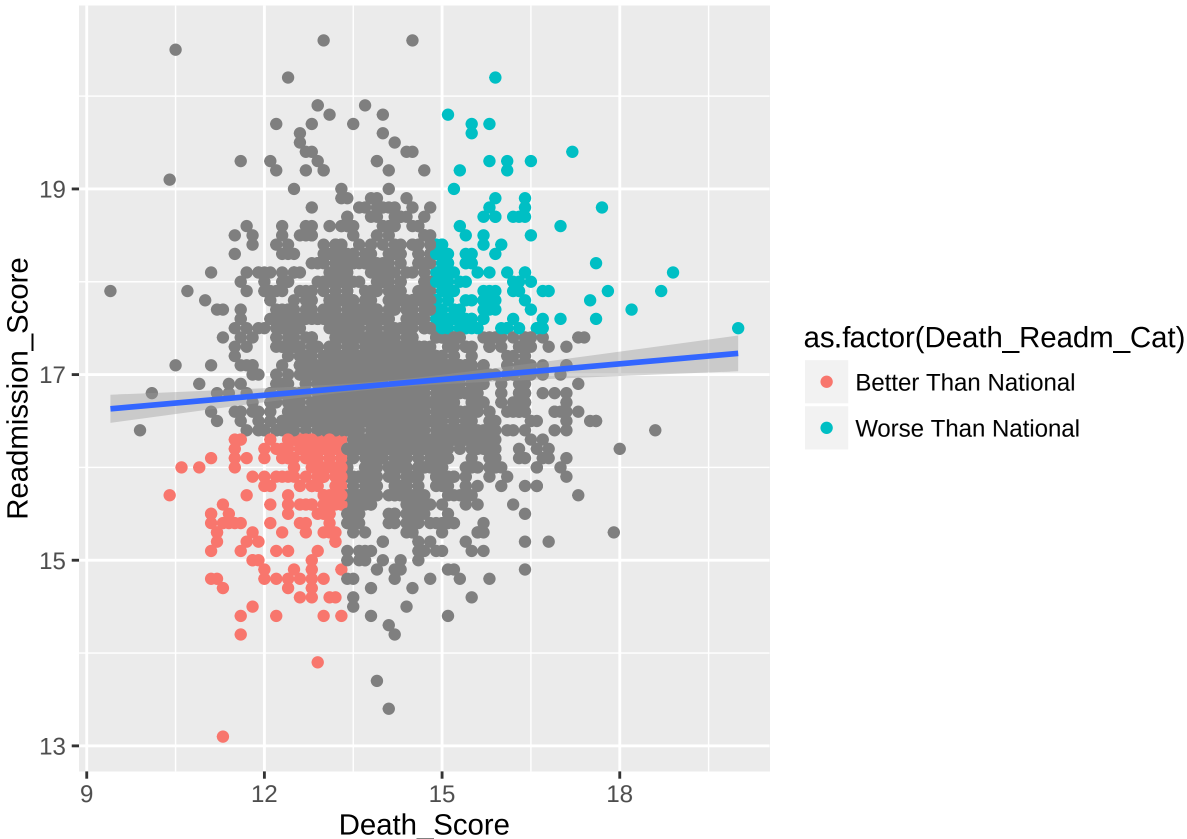
0.5
0.4
0.3
0.2
0.1
0.0
density
13
15
17
19
Readmission_Score


Mortality Rate
20
18
16
14
12
10


Readmission Rate
20
18
16
14



Average_Readm_Score
18
16
14
12
14
16
18
Average_Mort_Score
as.factor(Death_Readm_Cat)
Better Than National
Worse Than National



Analysis

Find best hospitals using 1st quartile, by hospital scores and by country average of scores, for mortaility and readmission Any hospitals below the 1st quantile, were considered one of the best hospitals. I did this for mortality and for readmission. This gave us the best hospitals according to mortality and readmission.

I then did this for counties (where I took the average of all hospital scores in that county) This gave us the best counties according to mortality and readmission.

I then found which of the best counties had at least one of the best hospitals in it.

I will quickly note that we live in one of the good counties, with Sarasota County having two hopsitals that are good according to this analysis with respect to both mortality and readmission.

Also both hospitals in the dataset for Linn, Iowa passed the good criteria for mortality and readmission, and Eau Claire, WI had 2 of 3 good for both criteria.

I also did a similar anaylsis for the worst counties and hospitals.

Both hospitals in the dataset for Leon, FL passed the bad criteria for mortality and readmission.

Both hospitals in the dataset for Ingham, Michigan passed the bad criteria for mortality and readmission.

Genesee, MI had 2 of 3 bad hospitals for both criteria.

This gave us a very conservative view of the best and worst counties in the nation.

All of the counties in this analysis have very few hospitals in them and they are either very small counties or very few of their hospitals were included in the dataset. For example, I know there are more than 4 hospitals in Sarasota County, but only 4 are in the dataset.

I decided that this would at best be a way to choose the best small county to go to, but this is problematic considering that not all hospitals are in the data and thus the total hospital count of a county is not a good indicator of how big or small the county is.

I also thought that judging these counties on such a small sample size of hospitals is not very fair, even if some of the conclusions are probably true.

I decided to widen the scope of the analysis and take a more liberal approach.

I noticed that taking aggragate averages of all hospitals in a county might way down counties that have really good hospitals, but just have a alot of average hospitals that crush these averages.

In addition, some counties have very good hospitals when considering their mortality rates, but might be considered bad with respect to readmission rates. I will look at these aspects separately in this analysis.

The remainder of the analysis will focus on the raw count and proportions of hospitals per county that are good or bad with respect to the 1st or 3rd quantiles.

This will bring some of the larger counties into the picture.

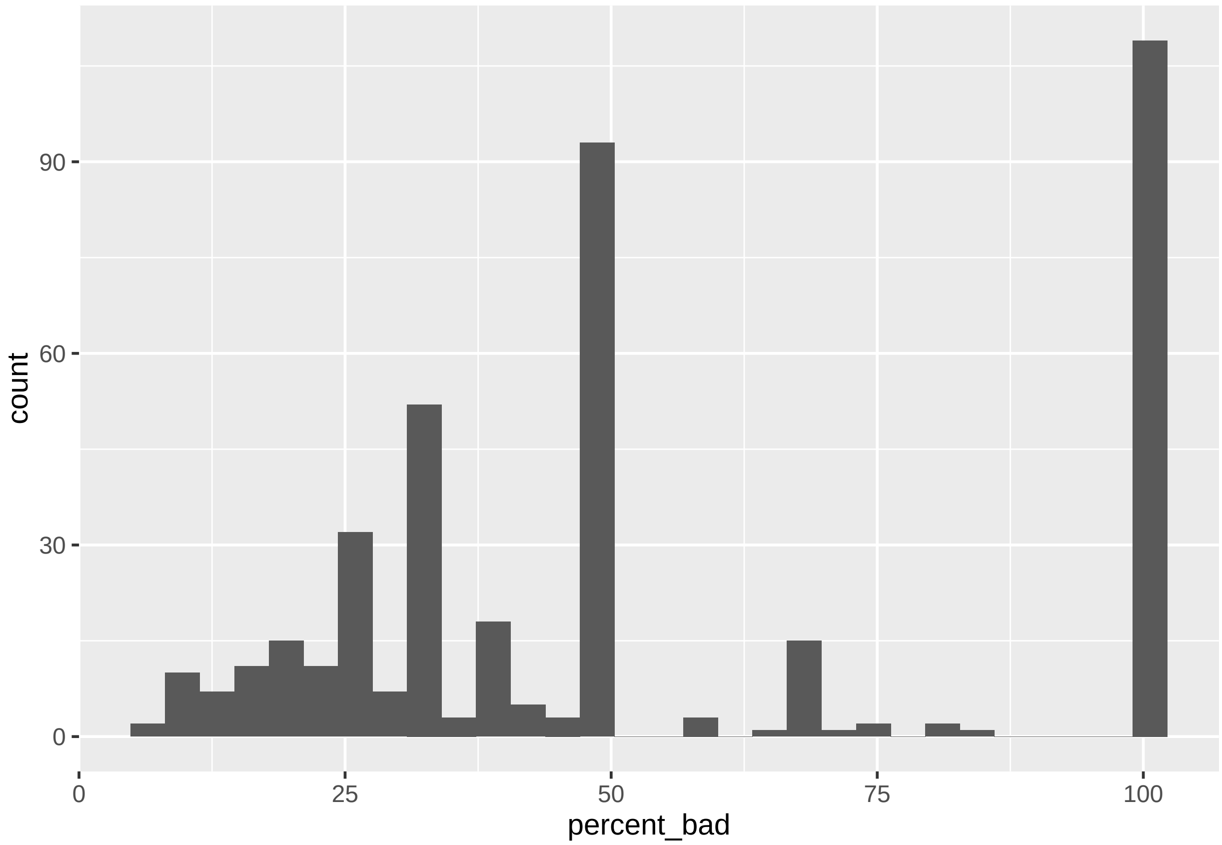
```
#Just by Mort
counties_best_mort <- hospitals_scores %>% filter(Death_Cat == "Better Than National") %>% group_by(Cou

counties_best_mort_avg <- inner_join(counties_best_mort, count_total_hosp, by = c("County", "State"))

counties_best_mort_avg$percent_bad <- counties_best_mort_avg$Count_Hospitals/counties_best_mort_avg$Cou

ggplot(counties_best_mort_avg, aes(percent_bad)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
counties_worst_mort <- hospitals_scores %>% filter(Death_Cat == "Worse Than National") %>% group_by(Cour

counties_worst_mort_avg <- inner_join(counties_worst_mort, count_total_hosp, by = c("County", "State"))

counties_worst_mort_avg$percent_bad <- counties_worst_mort_avg$Count_Hospitals/counties_worst_mort_avg$C

#ggplot(counties_worst_mort_avg, aes(percent_bad)) + geom_histogram()
head(counties_worst_mort_avg, n = 20)
```

```
## Source: local data frame [20 x 5]
## Groups: County [20]
##
##              County  State Count_Hospitals Count_Total_Hospitals
##              <fctr> <fctr>           <int>                 <int>
## 1             CLARK     NV               9                    17
## 2            DALLAS     TX               7                    29
## 3            HARRIS     TX               7                    40
## 4         RIVERSIDE     CA               6                    16
## 5              KING     WA               5                    17
## 6           TARRANT     TX               5                    22
## 7           BREVARD     FL               4                     7
## 8          MARICOPA     AZ               4                    37
## 9            MARION     IN               4                     9
## 10       MIAMI-DADE     FL               4                    21
## 11       PALM BEACH     FL               4                    13
```

```
## 12           QUEENS     NY               4                     8
## 13        SAN DIEGO     CA               4                    16
## 14        ALLEGHENY     PA               3                    18
## 15             COOK     IL               3                    49
## 16         DAVIDSON     TN               3                     9
## 17            DUVAL     FL               3                     7
## 18 EAST BATON ROUGE     LA               3                     7
## 19          EL PASO     TX               3                     8
## 20         ESCAMBIA     FL               3                     3
## # ... with 1 more variables: percent_bad <dbl>
```

```r
counties_best_mort2 <- counties_avg_scores %>% filter(Death_Cat == "Better Than National")
counties_worst_mort2 <- counties_avg_scores %>% filter(Death_Cat == "Worse Than National")
```

I now look at only those hosptials which are considered good with respect to mortality rate and took counts of these hospitals for each county. As far as raw numbers go, Los Angeles, CA has by far the most hospitals (31) that are considered good according to mortality rate. However the proportion of good to total hospitals is not very promising. Only 37% of the total hospitals are considered good by these standards. The first city that really jumps out with a good number of low mortality hospitals and also a good percentage is Suffolk County, NY with 9 of 11 hospitals considered good with respect to mortality. Providence, RI is also solid county with 6 of 7 hospitals in the dataset considered good in our analysis.

I also performed a similar analysis to find the worst hospitals with respect to mortality rate. Clark, NV has the highest number of hospitals with bad hospitals for mortality rate. They encompass 52% of the total hospitals in the county. There is a reasonably large number of total hospitals (17), so I am am tempted to say that this is the worst county to go to a hospital for heart attacks.

```r
# Just by Readm
counties_best_readm <- hospitals_scores %>% filter(Readm_Cat == "Better Than National") %>% group_by(Co

counties_best_readm_avg <- inner_join(counties_best_readm, count_total_hosp, by = c("County", "State"))

counties_best_readm_avg$avg <- counties_best_readm_avg$Count_Hospitals/counties_best_readm_avg$Count_To

counties_worst_readm <- hospitals_scores %>% filter(Readm_Cat == "Worse Than National") %>% group_by(Co

counties_worst_readm_avg <- inner_join(counties_worst_readm, count_total_hosp, by = c("County", "State"

counties_worst_readm_avg$avg <- counties_worst_readm_avg$Count_Hospitals/counties_worst_readm_avg$Count_

counties_best_readm2 <- counties_avg_scores %>% filter(Readm_Cat == "Better Than National")
counties_worst_readm2 <- counties_avg_scores %>% filter(Readm_Cat == "Worse Than National")
```

I then did the same analysis for readmission rate.

San Diego County, CA was the only county with 10 or more total hospitals in the dataset and also 50 % or more of them being condsidered good with respect to readmission rate.

Since there are no other counties that have high numbers of hospitals and large percentages of good ones for readmission rate, I will note that of the low count hospital counties, Sarasota has 3 of 4 hospitals with solid readmission rates.

When looking at the bad hospitals for readmission rates, Clark, NV was once again top 3 in bad hospitals (47% of total)