```
#HDFS

tail -n +2 care_hospital.csv > care_hospital_nohead.csv

hdfs dfs -mkdir care_hospital

hdfs dfs -put care_hospital_nohead.csv care_hospital

#Create table in HIVE

create external table care_hospital
(
id int,
name string,
address string,
city string,
state string,
zip int,
county string,
phone int,
condition string,
measure_id string,
measure_name string,
score string,
sample string,
footnote string,
start string,
end_date string
)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
 WITH SERDEPROPERTIES (
        "separatorChar" = ",",
        "quoteChar"     = '"',
        "escapeChar"    = "\\"
)

STORED AS TEXTFILE
LOCATION '/user/w205/care_hospital'
;

#Transform table to remove bad characters

CREATE TABLE effective_care as
SELECT care_hospital.condition, care_hospital.score_int FROM
care_hospital
WHERE care_hospital.score_int <> "High (40,000 - 59,999
patients annually)" AND
care_hospital.score_int <> "Low (0 - 19,999 patients
annually)" AND
```

```
care_hospital.score_int <> "Medium (20,000 - 39,999 patients
annually)" AND
care_hospital.score_int <> "Not Available" AND
care_hospital.score_int <> "Very High (60,000+ patients
annually)"
;

#Create another transform table

CREATE TABLE range as
SELECT condition,
max(score_int) as max,
min(score_int) as min
FROM effective_care
GROUP BY condition;

#Get final variability measure:

SELECT condition, max, min, max-min as difference
FROM range
ORDER BY difference DESC
;
```