```
# To install packages try: Pkg.add(["DataFrames", "Queryverse",
"SQLite", "IJulia", "Statistics"])
```

Objective

To rank the violation rate of doctors based on the incidence of
surgery for benign polyps.

Data

Medical data was was uploaded from SQLite Database provided by
Sarner. Diagnosis ICD-9-CM (D variables) and ICD-9-PCS (P
variables) were combined into single array strings D and P. Other
important variables like doctor_id and procedure were also
included.

```
# Make database connection
db = SQLite.DB("claims.db")

# Define a query to pull data drom the file claims.db
query =
"""
SELECT DISTINCT h.encounter_key, patient_id, doctor_id,
hospital_id, l.procedure,
(DA || ' ' || D1 || ' ' || D2 || ' ' || D3 || ' ' || D4 || ' ' ||
D5 || ' ' ||
D6 || ' ' || D7 || ' ' || D8 || ' ' || D9 || ' ' || D10 || ' ' ||
D11 || ' ' || D12 || ' ' || D13 || ' ' || D14 || ' ' || D15 || ' '
||
D16 || ' ' || D17 || ' ' || D18 || ' ' || D19 || ' ' || D20 || ' '
||
D21 || ' ' || D22 || ' ' || D23 || ' ' || D24 || ' ' || D25
) AS D,

(P1 || ' ' || P2 || ' ' || P3 || ' ' || P4 || ' ' || P5 || ' ' ||
P6 || ' ' || P7 || ' ' || P8 || ' ' || P9 || ' ' || P10 || ' ' ||
P11 || ' ' || P12 || ' ' || P13 || ' ' || P14 || ' ' || P15
) AS P

FROM medical_headers AS h
```