**My Project on Arrhythmia Recognition**

I selected a project on Arrhythmia Recognition because I ran into someone who sits for hours watching several screens of arrhythmia data at the hospital on Highway 3. It just seemed like have a ML algorithm help him do it would make his job so much easier.

I selected the data by looking at 5 datasets on line, narrowed them down to 2 then selected one

1) http://mlr.cs.umass.edu/ml/datasets/Arrhythmia

2) https://www.kaggle.com/bulentesen/cardiac-arrhythmia-database

3) https://www.kaggle.com/taejoongyoon/mitbit-arrhythmia-database

4) https://www.kaggle.com/alexandrefarb/mitbih-arrhythmia-database-de-chazal-class-labels#DS2\_signals.csv

5) https://www.kaggle.com/ronitf/heart-disease-uci

I have a Jupyter notebook where I examine these datasets (ExamineDatasets.ipynb)

I also had to read about arrhythmias so I could understand the data – primarily the EKG data (Arrythmia Notes)

And I talked to a patient care technician who has been reading EKGs for a few years. He gave me insight into the human part of this and how complex it can be.

(See section on Talking to a Tech and What to Look For in Arrythmia Notes)

Alex Stone from UTMB – highway 3 – has been monitoring for a couple years,

12 hours shifts and monitor anywhere from 3-35 people - but take breaks – several people monitoring in the same room, can cover for each other

Details about what he is looking for

QRS – infllunction/deflunction of your heart – when it turns into a U ­- depends on peak of QRS -happens quick – indication of heart attack.

Flag a J wave

the T wave -if it is prolonged – that means some type of heart damagae – bump upwards or downwards

Multiple T waves – not u waves – could be into afib

Multiple P waves – heart block

2 semesters of learning guidelines for reading EKGs

When a person moves the EKGs will spike and technicians can tell by how long they spike, what they know about the patient, etc.

to help them determine if it is just the person moving around, or if it really needs to be looked at

The role of sloppy data – 5 leads or 12 leads attached – issue with how good the data is

I read about how the electrical signal in the heart works – normally starts in one point and spreads across the heart before another one starts…

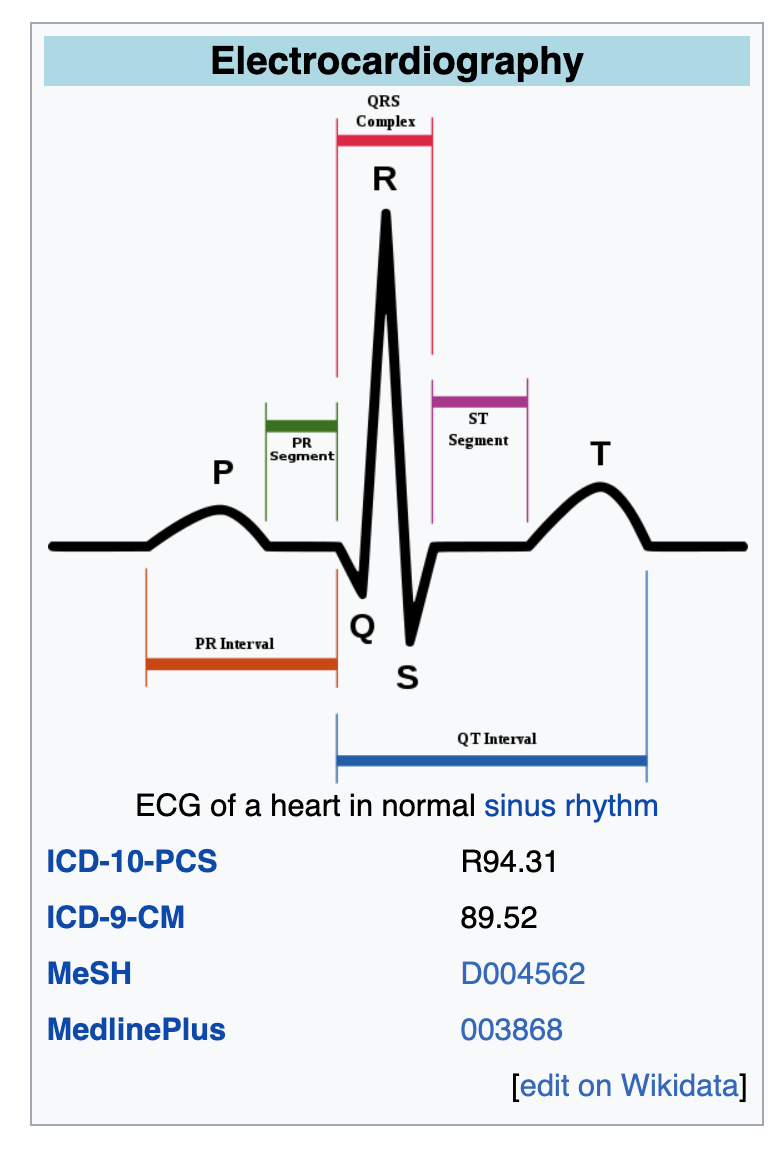
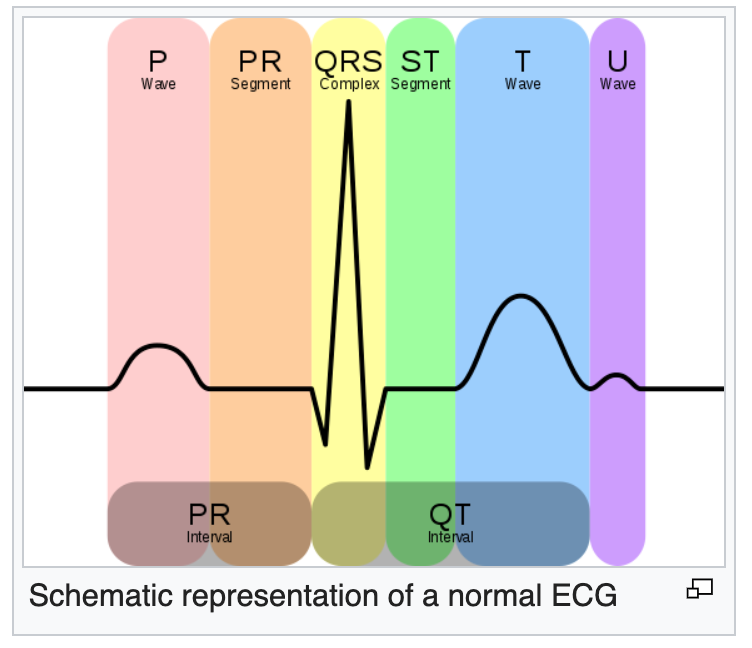
If another impulse starts – then the electrical signal is more complex

These waves are continuous and there are 5-12 leads per person.

My data is one row per EKG picture. Where they break it out into several characteristics (much like digit recognition), hence the large feature set size (280 features)

**Understanding the data:**

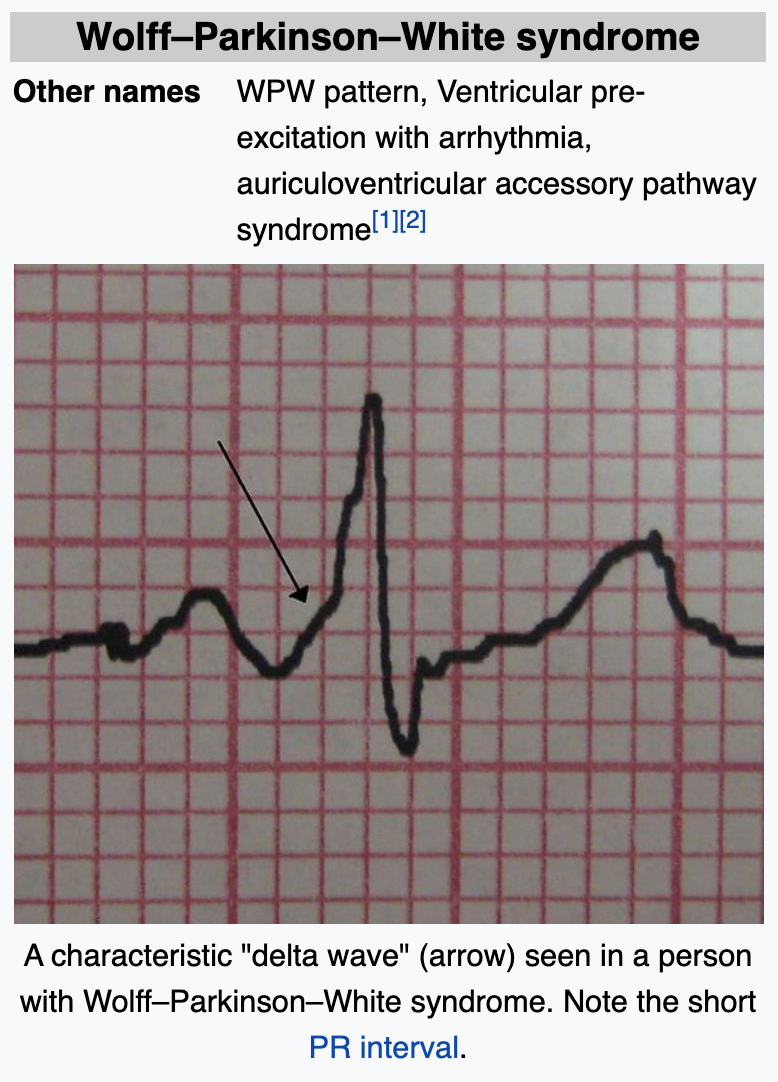
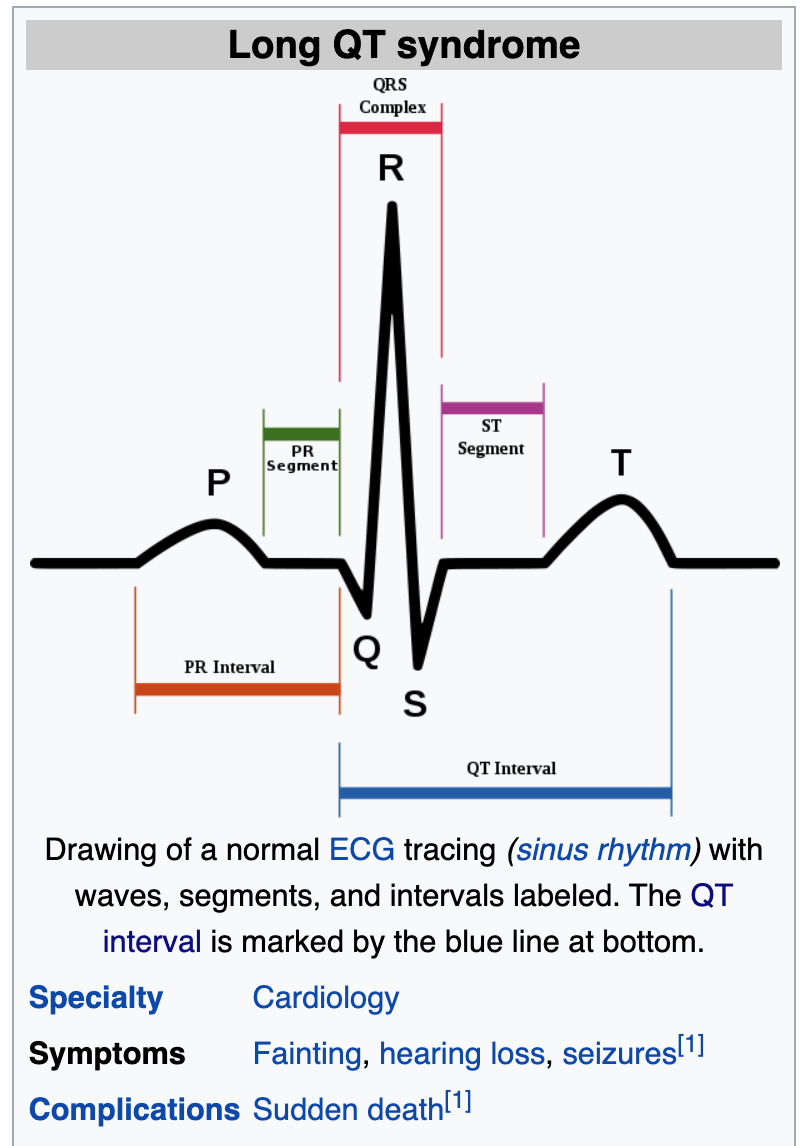
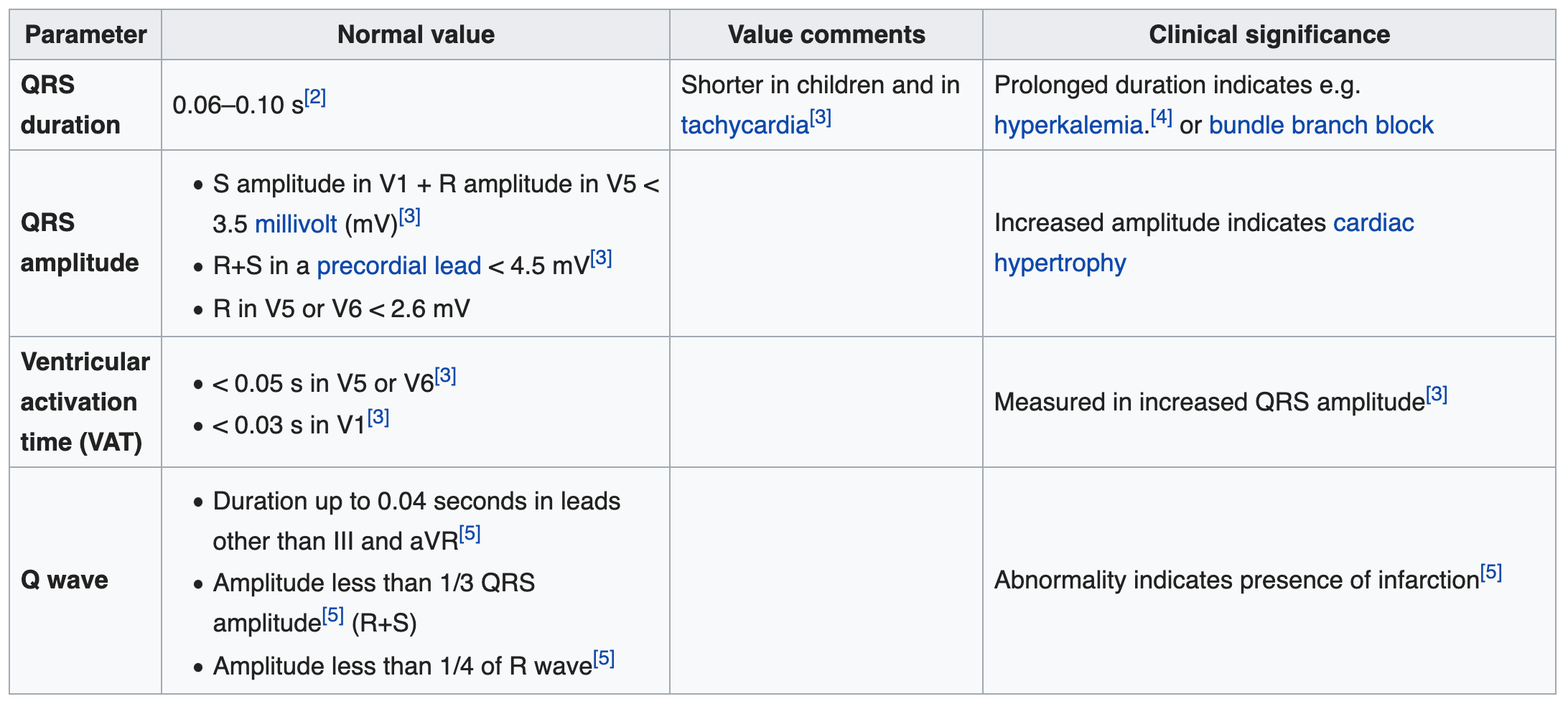
* The P wave represents atrial depolarization.
* The QRS complex represents ventricular depolarization.
* The T wave represents ventricular repolarization.
* The U wave represents papillary muscle repolarization.

The orderly pattern of depolarization (SA node, AV node, ventricles) gives rise to the characteristic ECG tracing.

ECGs can be recorded as short intermittent tracings (usually 10 second) or *continuous* ECG monitoring

**Examples of EKGs and what they mean:**

**Common Features in Arrhythmia datasets:**

The main thing in common is the pattern of the EKG broken up into the 5 parts -PQRST. There are also some J, U, R’\_wave, S’\_wave etc. but those are not as prevelant in the data.

Most data contained: age, sex, height, weight, heart rate, then the pattern of the EKG broken out, ex: QRS duration, PR interval, QT interval,T iterval, P interval, T, P, J, q\_wave, r\_wave, s\_wave, r’wave ,and s’wave.

**UNDERSTANDING THE SAMPLE SETS:   
arrhythmia.csv from http://mlr.cs.umass.edu/ml/datasets/Arrhythmia**

Sample of Arrhythmia.csv:

Number of Instances: 452

Number of Attributes: 279

? – means missing data

Last number is target .e.g. 8 represents an arrhythmia case

Concerning the study of H. Altay Guvenir: "The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. Class 01 refers to 'normal' ECG classes 02 to 15 refers to different classes of arrhythmia and class 16 refers to the rest of unclassified ones. For the time being, there exists a computer program that makes such a classification. However there are differences between the cardiolog's and the programs classification. Taking the cardiolog's as a gold standard we aim to minimise this difference by means of machine learning tools."

Attribute Information:

1-4:

75 age

0 male

190 height (cm) = 6.2 feet

80 weight (kilogram) = 176 lbs,

5-9:

91 (msec) avg QRS duration,

193 (msec) avg duration between onset of P and Q waves,

371 (msec) avg duration between onset of Q and T waves,

174 (msec) avg duration of T wave,

121 (msec) avg duration of P wave,

10-14:

-16 (degrees) avg angle of QRS wave,

13 (degrees) avg angle of T wave,

64 (degrees) avg angle of P wave,

-2 (degrees) avg angle of QRST wave,

? (degrees) avg angle of J wave,

15: 63 number of beats per min,

16-27: \*\*\* ON DI CHANNEL \*\*\* Then same but on channel DII, DIII, AVR, AVL, AVF, V1, V2, V3, V4, V5, V6 (LEADS)

0 (msec) width of Q wave,

52 (msec) width of R wave,

44 (msec) width of S wave,

0 (msec) width of R' wave,

0 (msec) width of S'wave,

32 number of intrinsic deflections,

0 Existence of ragged R wave, nominal

0 Existence of diphasic derivation of R wave, nominal

0 Existence of ragged P wave, nominal

0 Existence of diphasic derivation of P wave, nominal

0 Existence of ragged T wave, nominal

0 Existence of diphasic derivation of T wave, nominal

0, 52, 44, 0, 0, 32, 0, 0, 0, 0, 0, 0, DI

0, 44, 20, 36, 0, 28, 0, 0, 0, 0, 0, 0, DII

52, 40, 0, 0, 0, 60, 0, 0, 0, 0, 0, 0, DIII

52, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, AVR

0, 56, 36, 0, 0, 32, 0, 0, 0, 0, 0, 0, AVL

48, 32, 0, 0, 0, 56, 0, 0, 0, 0, 0, 0, AVF

80, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, V1

0, 40, 52, 0, 0, 28, 0, 0, 0, 0, 0, 0, V2

0, 48, 48, 0, 0, 32, 0, 0, 0, 0, 0, 0, V3

0, 52, 52, 0, 0, 36, 0, 0, 0, 0, 0, 0, V4

0, 52, 48, 0, 0, 32, 0, 0, 0, 0, 0, 0, V5

0, 56, 44, 0, 0, 32, 0, 0, 0, 0, 0, 0, V6

160-169: \*\*\* ON DI CHANNEL \*\*\* Then same but on channel DII, DIII, AVR, AVL, AVF, V1, V2, V3, V4, V5, V6 (LEADS)

-0.2 Amplitude , \* 0.1 milivolt, of JJ wave, linear

0.0 Amplitude , \* 0.1 milivolt, of Q wave, linear

6.1 Amplitude , \* 0.1 milivolt, of R wave, linear

-1.0 Amplitude , \* 0.1 milivolt, of S wave, linear

0.0 Amplitude , \* 0.1 milivolt, of R' wave, linear

0.0 Amplitude , \* 0.1 milivolt, of S' wave, linear

0.6Amplitude , \* 0.1 milivolt, of P wave, linear

2.1 Amplitude , \* 0.1 milivolt, of T wave, linear

13.6 QRSA , Sum of areas of all segments divided by 10, ( Area= width \* height / 2 ), linear

30.8 QRSTA = QRSA + 0.5 \* width of T wave \* 0.1 \* height of T wave. (If T is diphasic then the bigger segment is considered), linear

-0.2, 0.0, 6.1, -1.0, 0.0, 0.0, 0.6, 2.1, 13.6, 30.8, DI

0.0, 0.0, 1.7, -1.0, 0.6, 0.0, 1.3, 1.5, 3.7, 14.5, DII

0.1, -5.2, 1.4, 0.0, 0.0, 0.0, 0.8, -0.6, -10.7, -15.6, DIII

0.4, -3.9, 0.0, 0.0, 0.0, 0.0, -0.8, -1.7, -10.1, -22.0, AVR

0.0, 0.0, 5.7, -1.0, 0.0, 0.0, -0.1, 1.2, 14.1, 22.5, AVL

0.0, -2.5, 0.8, 0.0, 0.0, 0.0, 1.0, 0.4, -4.8, -2.7, AVF

0.1, -6.0, 0.0, 0.0, 0.0, 0.0, -0.8, -0.6, -24.0, -29.7, V1

0.0, 0.0, 2.0, -6.4, 0.0, 0.0, 0.2, 2.9, -12.6, 15.2, V2

-0.1, 0.0, 8.4,-10.0, 0.0, 0.0, 0.6, 5.9, -3.9, 52.7, V3

-0.3, 0.0, 15.2, -8.4, 0.0, 0.0, 0.9, 5.1, 17.7, 70.7, V4

-0.4, 0.0, 13.5, -4.0, 0.0, 0.0, 0.9, 3.9, 25.5, 62.9, V5

-0.3, 0.0, 9.0, -0.9, 0.0, 0.0, 0.9, 2.9, 23.3, 49.4, V6

DIAGNOSIS: 8

**DATASET #2**

**data\_arrhythmia.csv from https://www.kaggle.com/bulentesen/cardiac-arrhythmia-database**

Sample of data\_arrhythmia.csv:

Number of Instances: 453

Number of Attributes: 280

Missing data

Features:

age

sex

height

weight

qrs\_duration

p-r\_interval

q-t\_interval

t\_interval

p\_interval

qrs

T

P

QRST

J

heart\_rate

q\_wave

r\_wave

s\_wave

R'\_wave

S'\_wave

AA

AB

AC

AD

AE

AF

AG

AH

AI

AJ

AK

AL

AM

AN

AO

AP

AR

AS

AT

AU

AV

AY

AZ

AB'

BB

BC

BD

BE

BF

BG

BH

BI

BJ

BK

BL

BM

BN

BO

BP

BR

BS

BT

BU

BV

BY

BZ

CA

CB

CC

CD

CE

Cf

CG

CH

CI

CJ

CK

CL

CM

CN

CO

CP

CR

CS

CT

CU

CV

CY

CZ

DA

DB

DC

DD

DE

DF

DG

DH

DI

DJ

DK

Diagnosis

Open data\_arrhythmia with Numbers (double click on it) to see it best. Also last entry in DataSet2 notebook.

DataSet2 notebook tells how I cleaned the data: missing data and columns of zero data – I just removed those.