

Problem Set #3

MACSS 30000

Luxi Han, 10449918

Problem 1

The following graph is the lifetime temperature scatter plot:

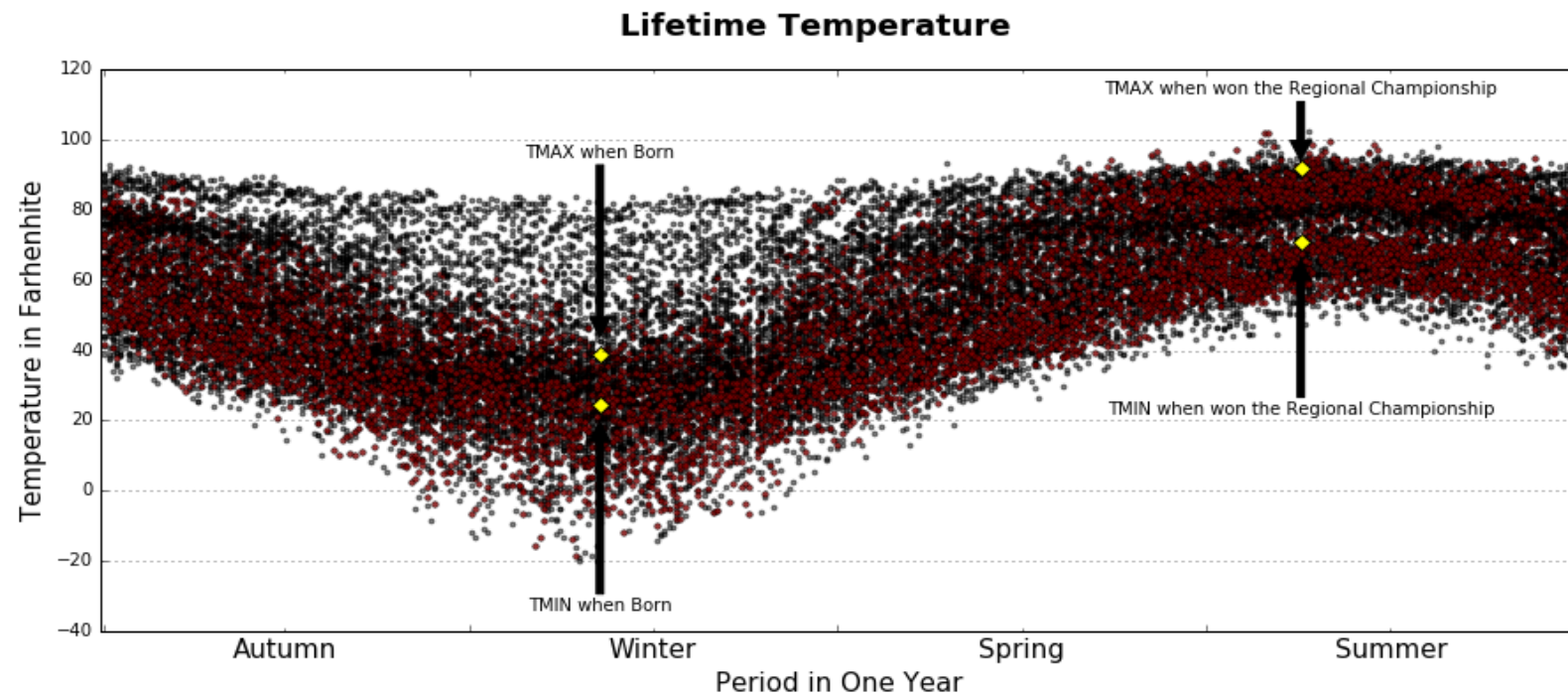


Figure 1: Lifetime Temperature Scatter Plot

Problem 2

(a)

Figure 2. shows the 2d histogram of cholesterol.
The midpoint of the highest frequency bin is 236.04.

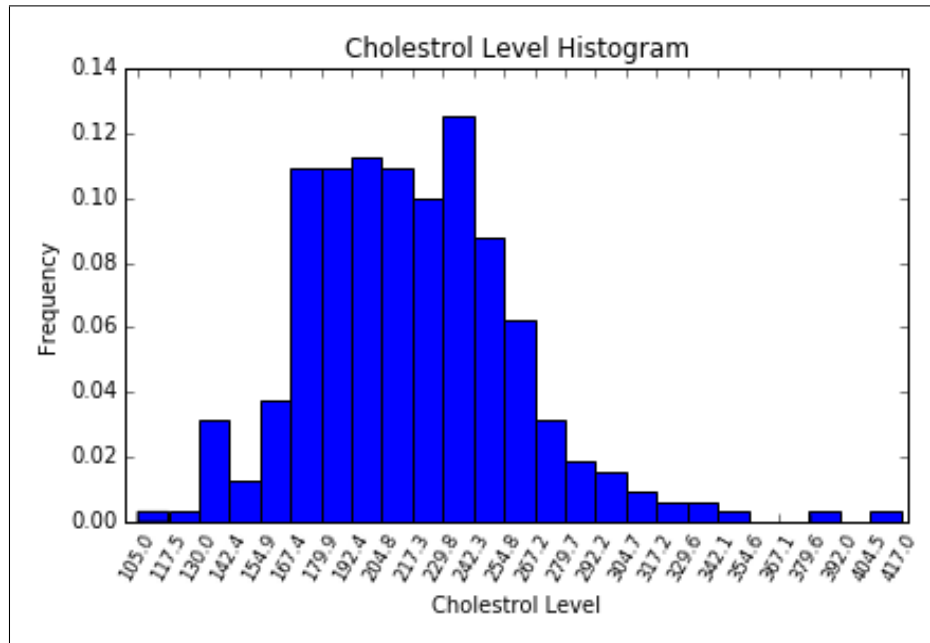


Figure 2: Histogram of Cholesterol Level

(b)

Figure 3. shows the 3d histogram of Cholesterol and Triglyceride Level. This adds the other axis which shows the number of people falling into different sub-groups of different Cholesterol and Triglyceride combination. It seems like most of the lipid patients have cholesterol level around 150 - 250; and the triglyceride level 70 - 180.

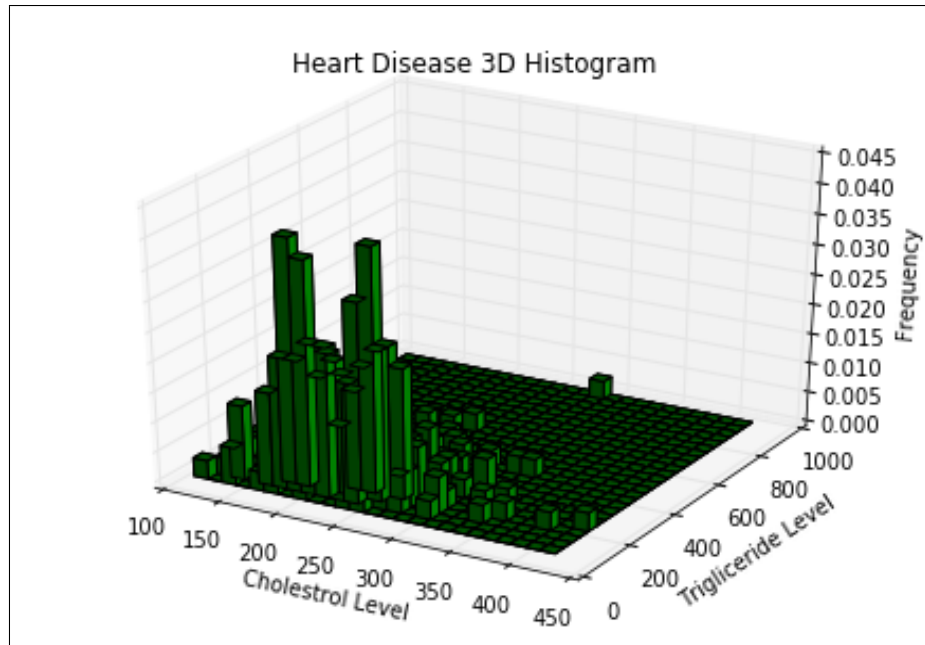


Figure 3: 3D Histogram of Cholesterol and Triglyceride Level

(c)

According to the histogram, we should deduce that the groups with the highest frequency are the ones that have the highest probability to have heart disease, which would mean the individuals with cholesterol level between 150 - 250 and with triglyceride level between 70 - 180.

But one thing worth mentioning is that **it can well be the case that higher cholesterol level and higher triglyceride level will lead to higher chance of having heart disease**. The reason why we don't observe this in the data set is simply because there aren't that many people with cholesterol and triglyceride level above a certain level. In other word, there's systematic sample selection bias in the data set.

Problem 3

Figure 4. shows the Recession Spider Graph of job employment level.

j) The 1945 recession is worse than the Great Recession at the first couple of month immediately following the peak month but then the job employment level returned to a much more robust level than the Great Recession.

k) In terms of job loss, the Great Recession is in general better than the Great Depression.

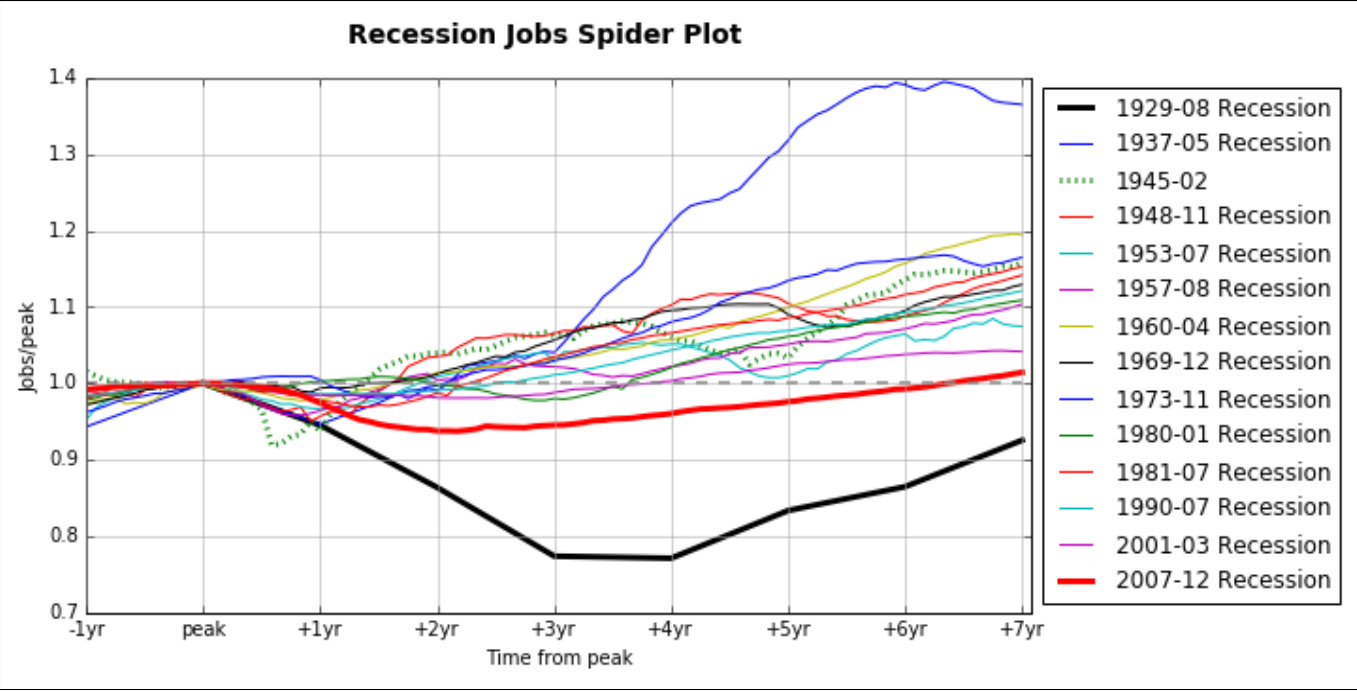


Figure 4: Spider Plot of Job Employment Level during Recession