Question 1
Coded with MySQL:
bughistory refers to the dataset BugHistory.csv, and bugs refer to Bugs.csv, with the columns named as given from the dataset.

```
1     /* Question 1 */
2     SELECT  bughistory.FullDate,
3             COUNT(bughistory.BugId) AS num_bugs
4     FROM bughistory
5     JOIN bugs
6     ON bugs.BugID = bughistory.BugId AND
7        bugs.Severity = 'A'
8     GROUP BY bughistory.FullDate
9     HAVING num_bugs > 7
10    ORDER BY num_bugs DESC;
```

Example Results:

FullDate	num_bugs
2017-07-11	34
2017-07-19	34
2017-07-12	34
2017-07-20	34
2017-07-13	34
2017-07-21	34
2017-07-14	34
2017-07-22	34
2017-07-15	34
2017-07-23	34

Qustion 2:

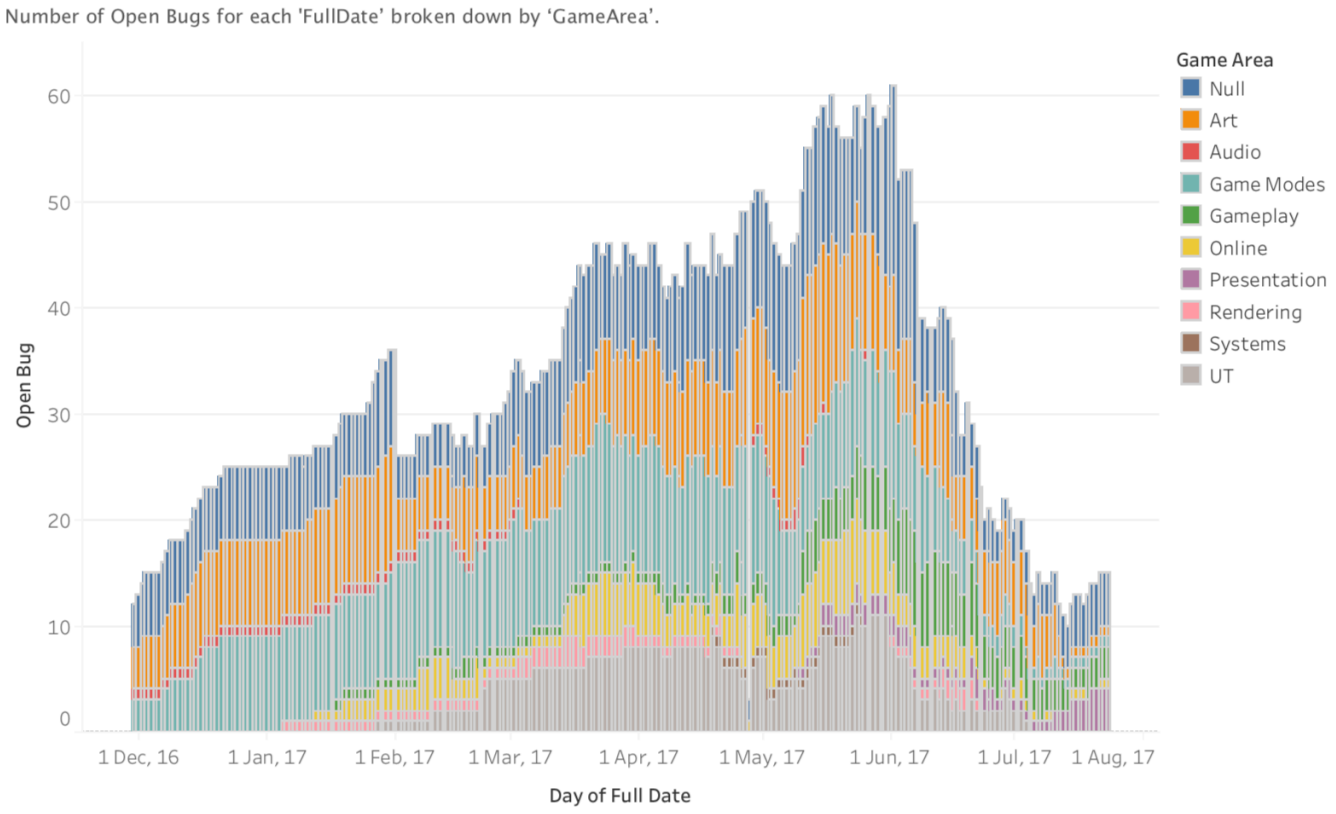
```
12    /* Question 2 */
13    /* Part A */
14    SELECT bugs.GameArea,
15           bughistory.FullDate,
16           COUNT(bughistory.ProgressStatus) AS open_bug
17    FROM bugs
18    JOIN bughistory
19    ON bugs.BugId = bughistory.BugID AND
20       bughistory.ProgressStatus IN ('Claim Fixed',
21                                     'Confirmed',
22                                     'Fix Failed',
23                                     'In Progress',
24                                     'Info Added',
25                                     'Pending Build',
26                                     'Request Review',
27                                     'Submitted')
28    GROUP BY FullDate, GameArea;
```

Part A example results:

	GameArea	FullDate	open_bug
▶		2016-11-29	4
		2016-11-30	4
		2016-12-01	4
		2016-12-02	5
		2016-12-03	5
		2016-12-04	5
		2016-12-05	5
		2016-12-06	5
		2016-12-07	5
		2016-12-08	5

Part B:

I chose to use Tableau to make a bar graph of the total amount of open bugs according to date, and broken down to different game areas with different colours.



Part C:

We can see the continuous growth trend of bug testing, and applying updates causing fluctuations or a decrease of open bugs in the graph, specifically end of Apr 2017, which might have been an major update for gameplay. Open bugs are focused more on Null, Art, UT, and Game Modes during production, and the amount of bugs in gameplay and presentation rises when reaching finalization, and this data can explain which game areas to focus during development process.

Question 3

a) The first measure that I would recommend to compare is the residual standard error, in where the smaller this number, the better the fit of the regression model, and the smaller amount of error in the least square line compared with the data points. Then another measure that is also very important is the R-squared and Adjusted R-squared value, the larger this value gets, means the smaller the sum of squares residual, and the better the fit of the model. Adjusted R-squared is better to use in comparison, as R-squared value always increases with additional explanatory variables. Overall, larger adjR-square and smaller residual leads to a better prediction equation.

b) - Our model is assumed to be a random variable with non-random explanatory variables, and the residuals in our model are independent normal random variables with a mean of 0 and constant variance.
   - The data should be randomly sampled, and the observations must be taken without error.
   - Their must be a linear relationship in the model, should check scatterplots for homoscedasicity or curvilinear relations.
   - No multicollinearity, which can be checked with VIF test or correlation matrix.

c) Severity3, largest t-value and smallest P-values indicates statistical significance. It has the largest absolute valued estimate, an unit increase in severity3 decreases BugWeight by 983.25.