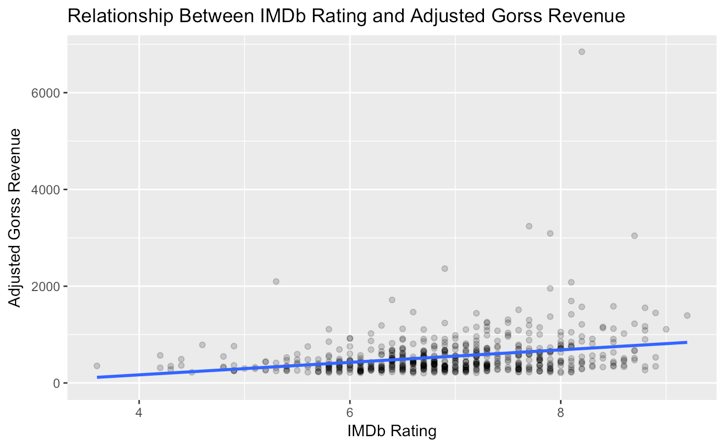
**Assignment 2-1**

Group assignment (20’)

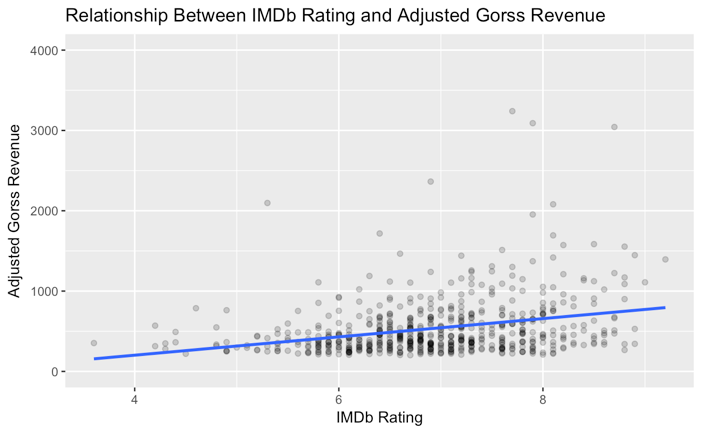
* Github collaboration (final codes) – due Nov 2
* Report – due Nov 2
* Presentation at next class
  + Think about who, what, and how
  + Think about your storyline
* Grade will be given after next class

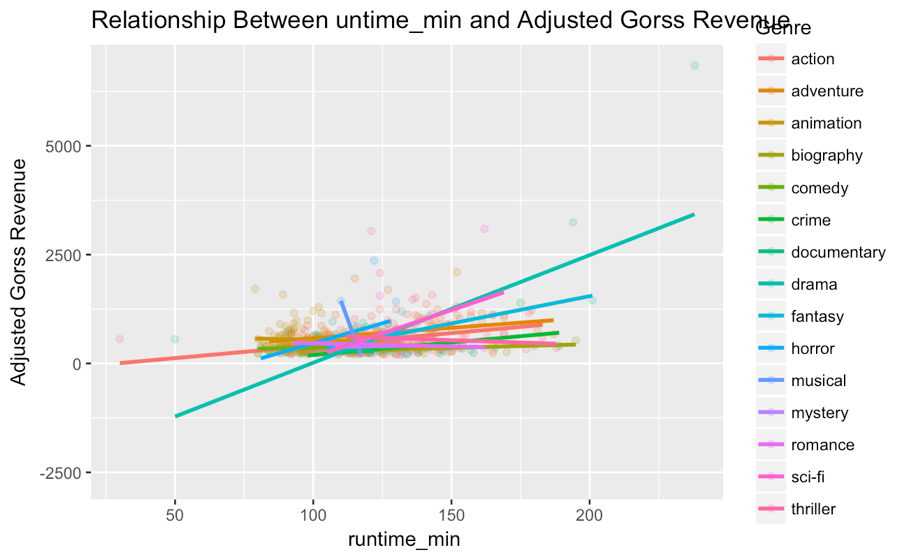
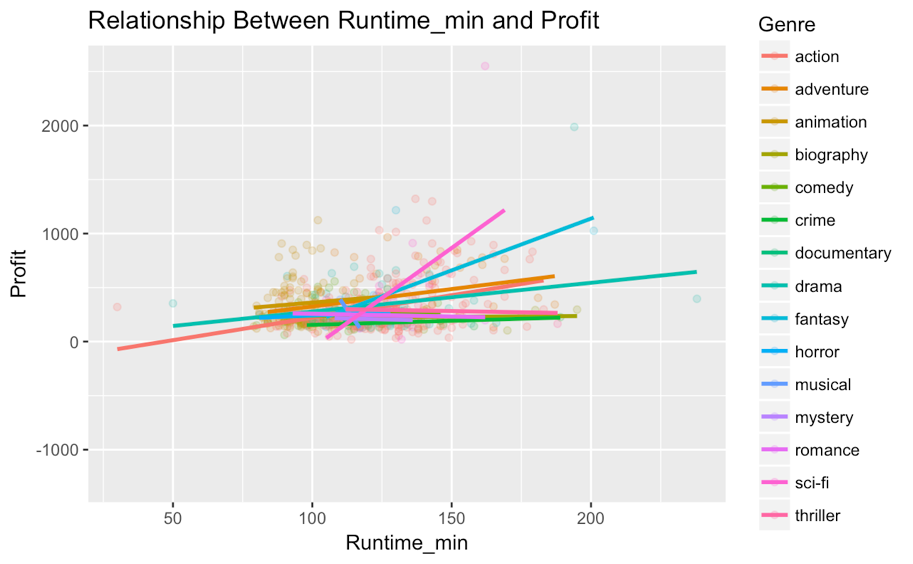
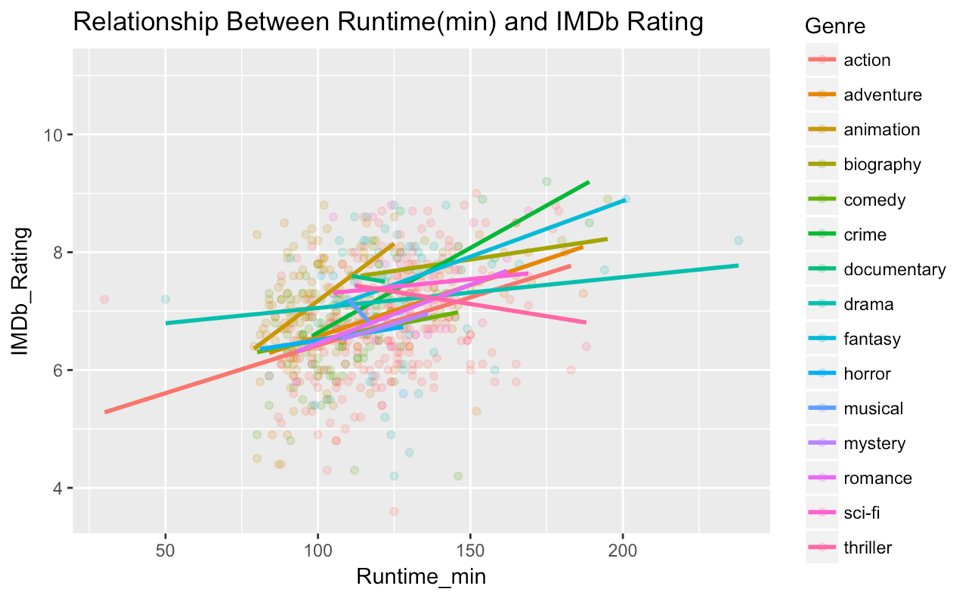
Netflix’s CEO is impressed with your initial findings. Now he asked a few follow-up questions

* **Does higher rating movie also have a higher adjusted gross revenue?**
  + - Given the IMDb Rating and MovieLense Rating is highly correlated, in order to avoid multicollinearity, we only adopt IMDb Rating as the measure of rating (Pearson coefficient = 0.88).
    - Intuitively speaking, the linear relationship between IMDb Rating and adjust gross revenue is not strong.

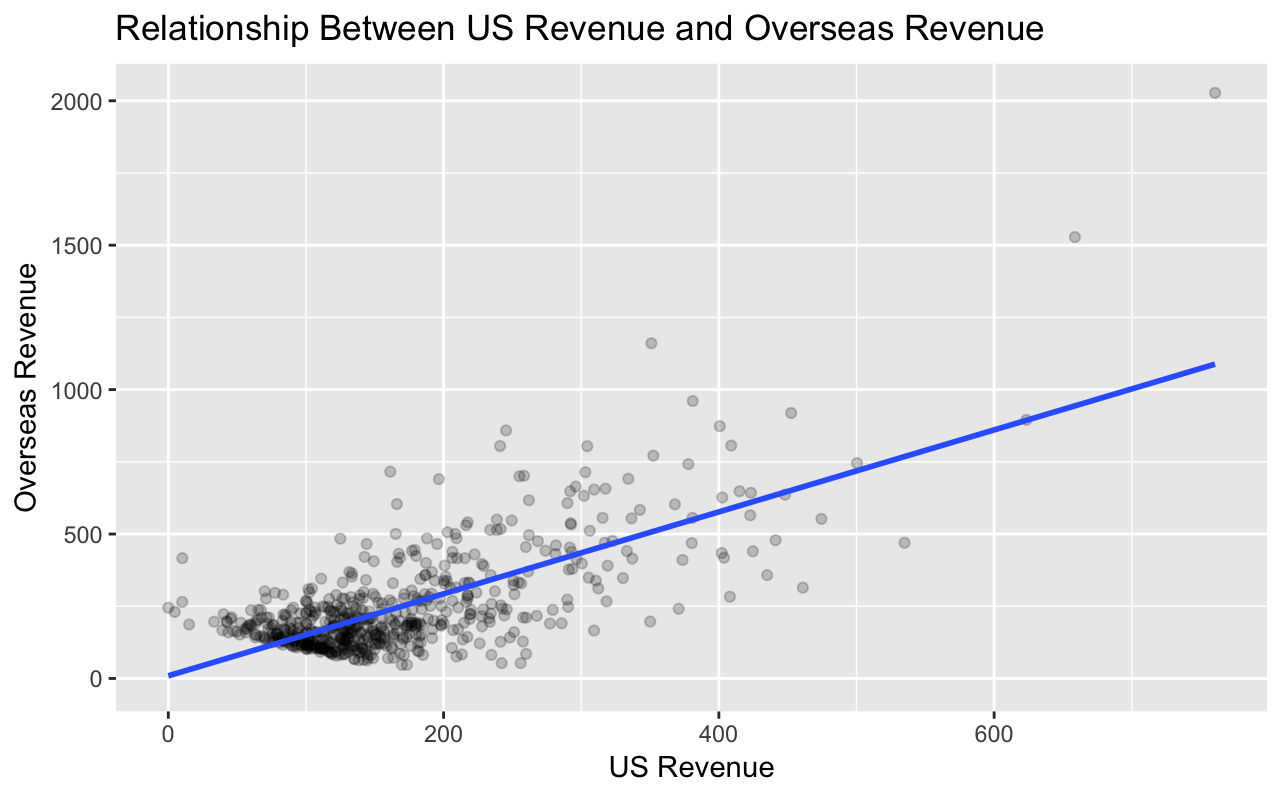
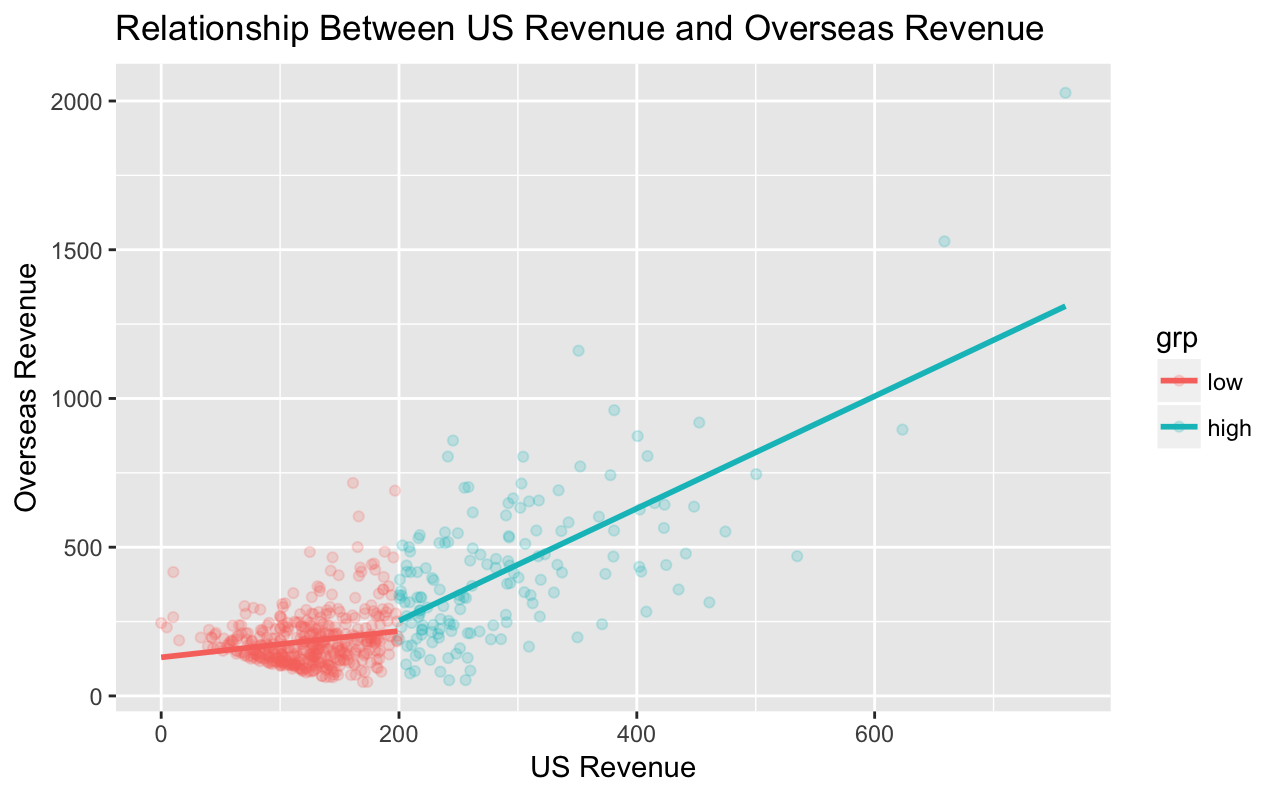


* Omit the outlier (with AGR>6000) – seems the outlier does not influence the trend significantly



* + - Pearson coefficient = 0.27
    - Higher rating movie might have a higher adjusted gross revenue, but the correlation relationship is not strong.
* **Shall we make a short movie or a long movie?**
  + It depends on the object – to maximize gross revenue? Maximize profit? To increase ratings (good reputation)?
  + It also depends on genre
    - Time & Gross Revenue
    - 
    - time & profit
    - 
    - time& rating
    - 
    - we can also compare the coefficients in a table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *genre\_name* | *time\_revenue* | *time\_rating* | *~~time\_profit~~* | *Type* |
| 1 | action | 0.36 | 0.36 | ~~0.37~~ | 1 |
| 2 | adventure | 0.29 | 0.54 | ~~0.34~~ | 1 |
| 3 | animation | -0.04 | 0.38 | ~~0.09~~ | 3 |
| 4 | biography | 0.42 | 0.38 | ~~0.05~~ | 1 |
| 5 | comedy | 0.16 | 0.20 | ~~0.11~~ | 1 |
| 6 | crime | 0.47 | 0.78 | ~~0.40~~ | 1 |
| 7 | documentary | -1.00 | -1.00 | ~~1.00~~ | 1 |
| 8 | drama | 0.66 | 0.12 | ~~0.25~~ | 1 |
| 9 | fantasy | 0.76 | 0.68 | ~~0.68~~ | 1 |
| 10 | horror | 0.43 | 0.16 | ~~0.25~~ | 1 |
| 11 | musical | -1.00 | -1.00 | ~~-1.00~~ | 1 |
| 12 | mystery | 0.06 | 0.95 | ~~-0.10~~ | 3 |
| 13 | romance | -0.21 | 0.67 | ~~-0.15~~ | 2 |
| 14 | sci-fi | 0.52 | 0.11 | ~~0.56~~ | 1 |
| 15 | thriller | -0.10 | -0.17 | ~~-0.07~~ | 1 |

* + - type1: For genres like action, adventure, crime etc, runtime has consistent effects (positive or negative) on both revenue and rating. The longer, the better.
    - Type2: But runtime has opposite effects on the revenue and rating for genres like romance. So whether longer time or short time is better depends on whether our priority- revenue or rating (reputation).
    - Type3: Runtime has very little correlation with revenue, but highly correlated with rating. (i.e. animation and mystery).
* **If a movie does well in US, does it also usually do well overseas?** 
  + Define ‘do well’ by the amount of revenue
  + Overall: there’s correlation relationship between US revenue and Overseas revenue. Nut seems US Revenue=200 is cutoff – within 0< US Revenue<200, data points are highly overlapping and gather below the lin. While data points within (200, inf) is more separated and evenly fall on both sides of the line.
  + 
  + if we use US Revenue=200 as a cutoff and separate the data points into 2 groups – low US Revenue and high US Revenue, we get:
  + 
    - Overall, if revenue in the US is positively correlated with revenue overseas (0.71).
    - US Revenue > 200: if the movie does well in the US, there’s high probability that the movie will also do well overseas (0.67).
    - 0 < US Revenue < 200: the US revenue is weakly correlated with the overseas revenue (0.19).
* Now you have explored many variables and how they affect revenue, please recommend a strategy for Netflix’s next investment in a movie.