1. What are the names of the people in your group?

Tim Abramov

1. What research topic or question are you interested in?

Analyzing and visualizing data about video games.

1. What type of project will this be? Will you be analyzing a data set, creating a visual narrative of data, using R to create simulations, etc.?

I plan to do most of data cleanup, analysis, and visualization using R.

1. Do you plan to collect your own data, or do you hope to use a data set that already exists (or is simulated)? If you plan to use a pre-existing data set, have you already found it?

Dataset was provided to me by a Professor.

1. Describe your likely data set.
   1. List the unit of study in your data. For example, if your data is in a spreadsheet, does each row represent one person (or one movie, or one country, etc.)?

Each row in the dataset represents an individual video game.

* 1. How big is the data? How many observations are there?

Dataset consists of 13,357 records.

* 1. List the variables you will measure or obtain data about, including units when appropriate.

There are too many variables to list them all, so I’ll list more notable ones: game name, release date, Metacritic score, owners count, price, genre, and platform.

1. Statistical questions: List several specific questions that can be answered using the variables in your data set.

* Not all games receive a Metacritic score, is there an association between how many people own the game and whether game has a Metacritic score? Also, would be interesting to see if any other variables might also have an effect.
* For games that already have a Metacritic score, can we find out what parameters influence that score the most (have the strongest association)?
* Is there an association between how many copies were sold and price of a game?
* Is there an association between game description length (in words) and price of the game or copies sold?

1. What is your background reading plan? Give a short list of topics you need to research for your project.

* Need to find out how to clean/pre-process the data in R and how it’s usually done
* Should I look at any research in this area?