Godin_review

Some comments on your repo are as follows:

- I feel the repository itself is quite messy. Getting rid of some of the auxiliary files or storing them in folders would be nice and help users of the repository. For example, you could store all the photos in the same folder.
  - Along these lines, I found that the organization of the output files was confusing. You have v1 and v2 files for the chromosomes, but it would be better if you named them as increasing or decreasing or put them in separate subfolders. A good rule of thumb is that visitors of your repo shouldn't have to guess what files are. They should have a good idea from the name, readme file, and directory structure (folder names).
  - Also, you shouldn't have version names in your photos as Git will track this. For example, you have two versions of the "homozygous proportion" file and I'm not sure if they are both relevant or if only one is.
- I could NOT get the code in your R markdown file to run. It doesn't appear that there are any code blocks. These are denoted by `{R} code` (see the working example below). I would add this to your markdown file so that it runs properly. I would also divide up the code into chunks, like you did, so that the user can run the chunks independently.
- For your decreasing file position folders, it doesn't appear that you replaced the missing data symbol of ? with - as requested. At least, it wasn't done so for maize_sortedv2_1.txt. One way to do this would be the gsub function in R: https://stackoverflow.com/questions/68586589/using-gsub-function. The assignment is NOT asking you to substitute the NA data with ? or -, it is asking you to replace entries like `?/?` with `-/-` for the actual SNP data.
- For your SNP per chromosome chart `SNP count per chromosome-V2.jpeg`, why do you have error bars? The are no replicate experiments where you would have variability in this.
- I didn't see any graph which looked at the distribution of SNPs across locations among the different chromosomes, which I think is what they were asking for when they asked for the distribution on chromosomes.
- In your picture looking at the homozygous distribution among groups, I think they meant to analyze for the groups like ZMMLR or TRIPS, not the different species maize and teosinte.

Working R examples.

```R
fang_data <- read.table("fang_et_al_genotypes.txt", header = TRUE, sep = "\t", stringsAsFactors = FALSE)
snp_data <- read.table("snp_position.txt", header = TRUE, sep = "\t", stringsAsFactors = FALSE)
```