R Notebook, Week of 8/18

Contents

MAMBA Simulations **2**

SNPs With Outliers 3

SNPs Without Outliers 6

Inverse Variance Weighted 9

Giant Consortium Studies **10**

BMI 10

Height 12

Prior Checking **13**

Meta Graphs 13

MAMBA Simulations

First we simulate data according to MAMBA and calculate the PPR values (from the MAMBA package) and the PRP values (from the PRP package). The job R files used for this are located in the code/data directories. The packages are calculated for all 50000 SNPS. Here, we'll look at the case where the nonoutlier study rate is 0.975.

```
load(file = "data/mamba_data/sim_mamba_mod_p975.rda")
load(file = "data/mamba_data/mamba_data_p975.rda")

pprs <- sim_mod$ppr
```

We take the indices for the SNPs that have at least one outlier study.

```
# indices for snps w/ and w/o outliers
out_studies <- mamba_data$Ojk
out_rows_ind <- which(rowSums(out_studies == 0) > 0) # indices of snps with outlier studies
no_out_rows_ind <-which(rowSums(out_studies) == 10) # indices of rows w/o outliers
```

Then we load in the data containing the PRPs (from PRP package) for each SNP.

```
# loading PRP data
load(file = "data/prp_data/post_prp_data_pval_p975.rda") # post_prp_data_pval
```

Below are the PPRs and PRPs from the the SNPs with and without outlier studies.

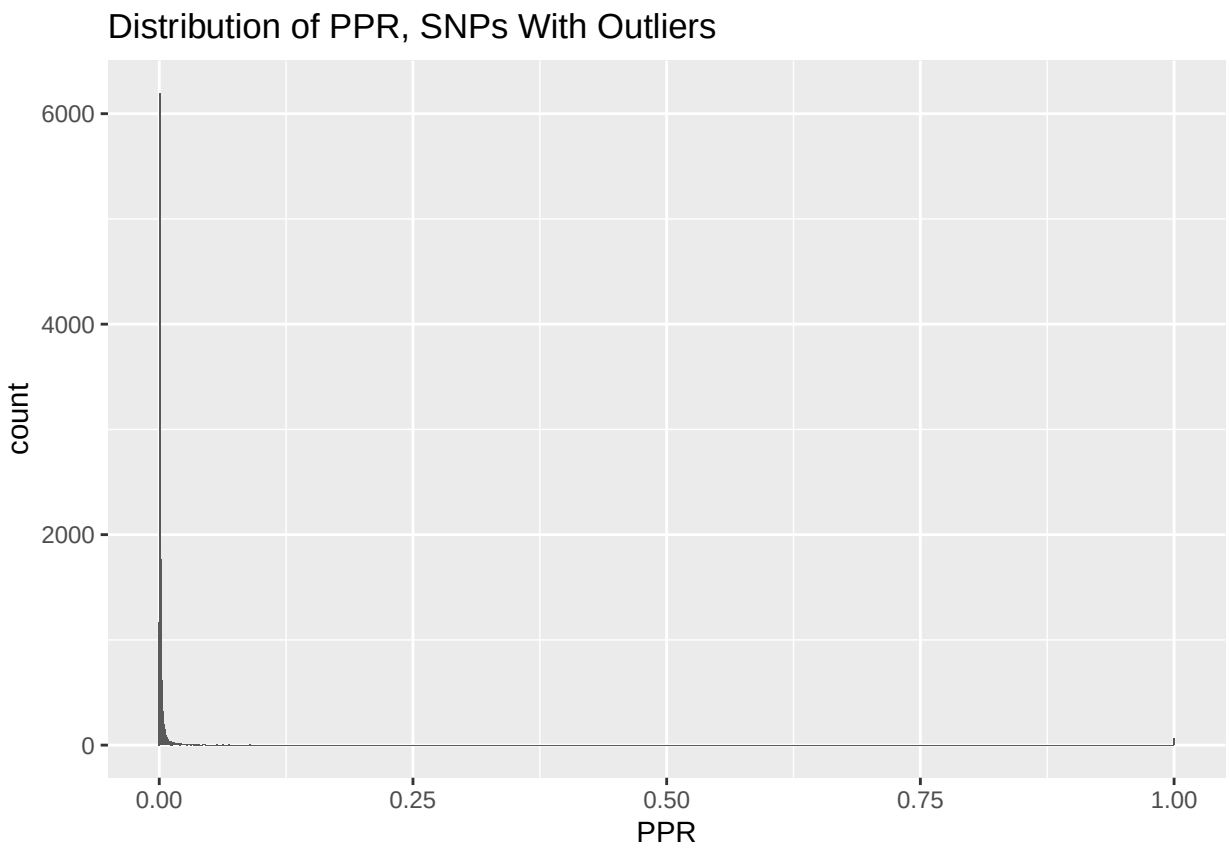
```
# MAMBA pprs for snps w/ and w/o outliers
out_ppr <- pprs[out_rows_ind]
nonout_ppr <- pprs[no_out_rows_ind]

# prps for snps w/ and w/o outliers
out_prp <- post_prp_data_pval[out_rows_ind]
nonout_prp <- post_prp_data_pval[no_out_rows_ind]
```

SNPs With Outliers

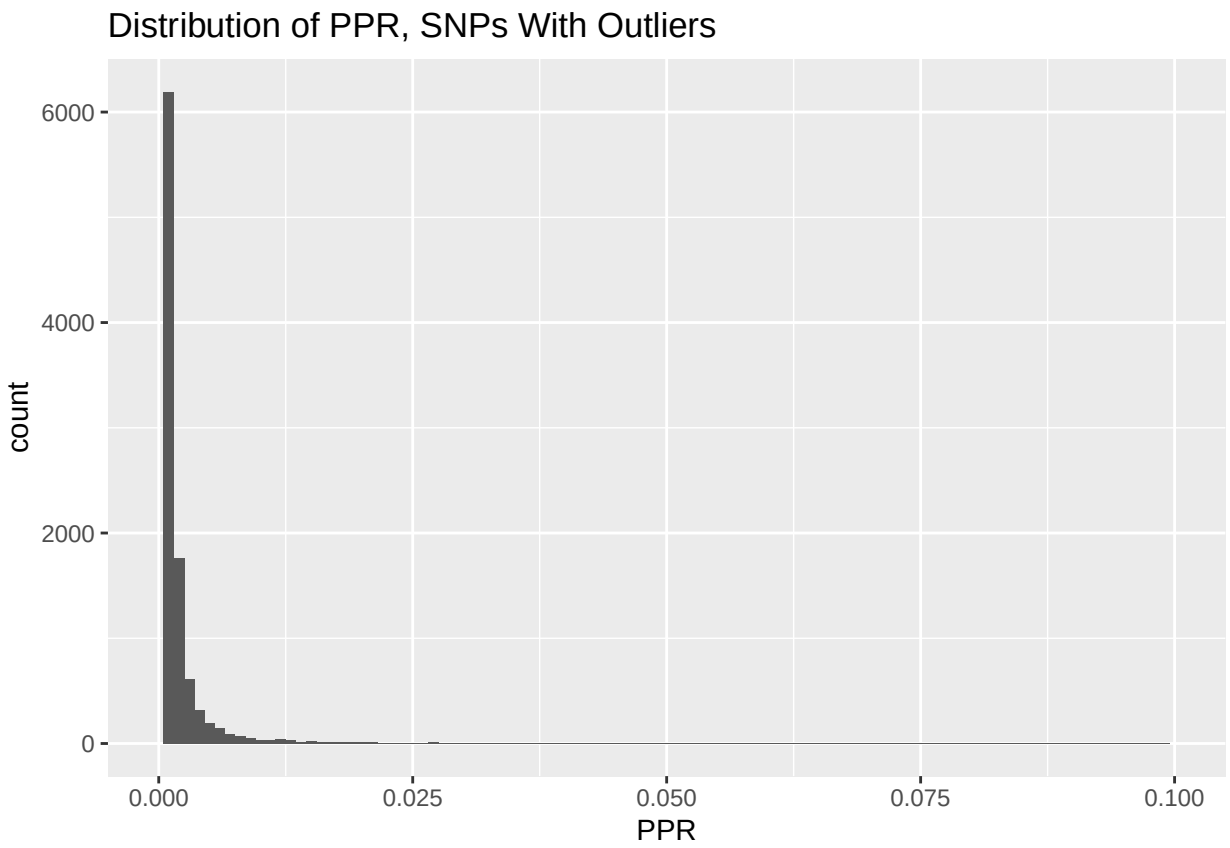
For our SNPs **with outliers**, the distribution of our PPRs looks as follows:

```
ggplot(data = as.data.frame(out_ppr), aes(out_ppr)) +
  geom_histogram(binwidth = 0.001) +
  ggtitle("Distribution of PPR, SNPs With Outliers") +
  xlab("PPR")
```



Rescaling the graph by ignoring some outliers, we get the histogram below.

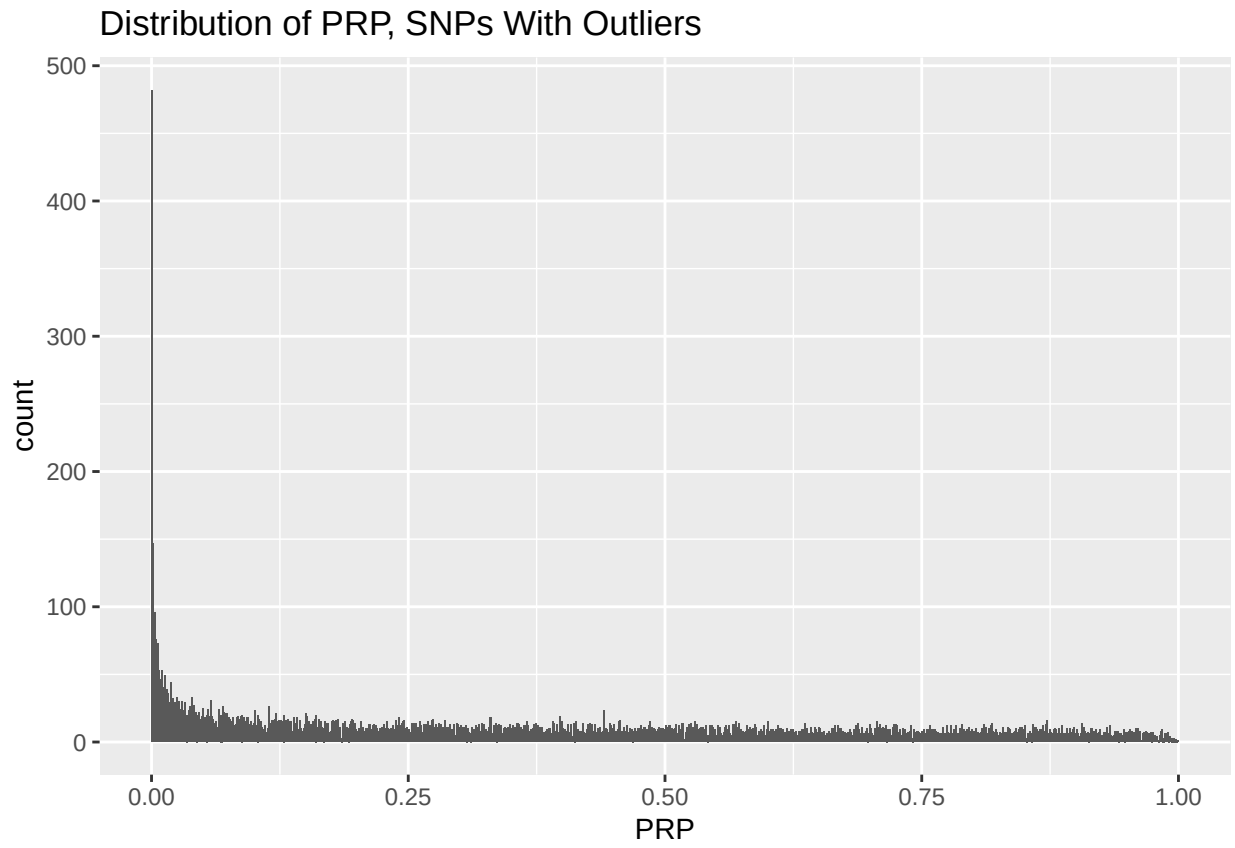
```
ggplot(data = as.data.frame(out_ppr), aes(out_ppr)) +
  geom_histogram(binwidth = 0.001) +
  ggtitle("Distribution of PPR, SNPs With Outliers") +
  xlab("PPR") +
  xlim(0, 0.1)
```



With outliers, there are around 11004 SNPS with PPR values less than or equal to 0.05 (our significant SNPs).

In contrast, this is what the distribution of our PRPs for outlier SNPs looks like. Keep in mind the differently scaled x-axis.

```
ggplot(data = as.data.frame(out_prp), aes(out_prp)) +
  geom_histogram(binwidth = 0.001) +
  ggtitle("Distribution of PRP, SNPs With Outliers") +
  xlab("PRP")
```

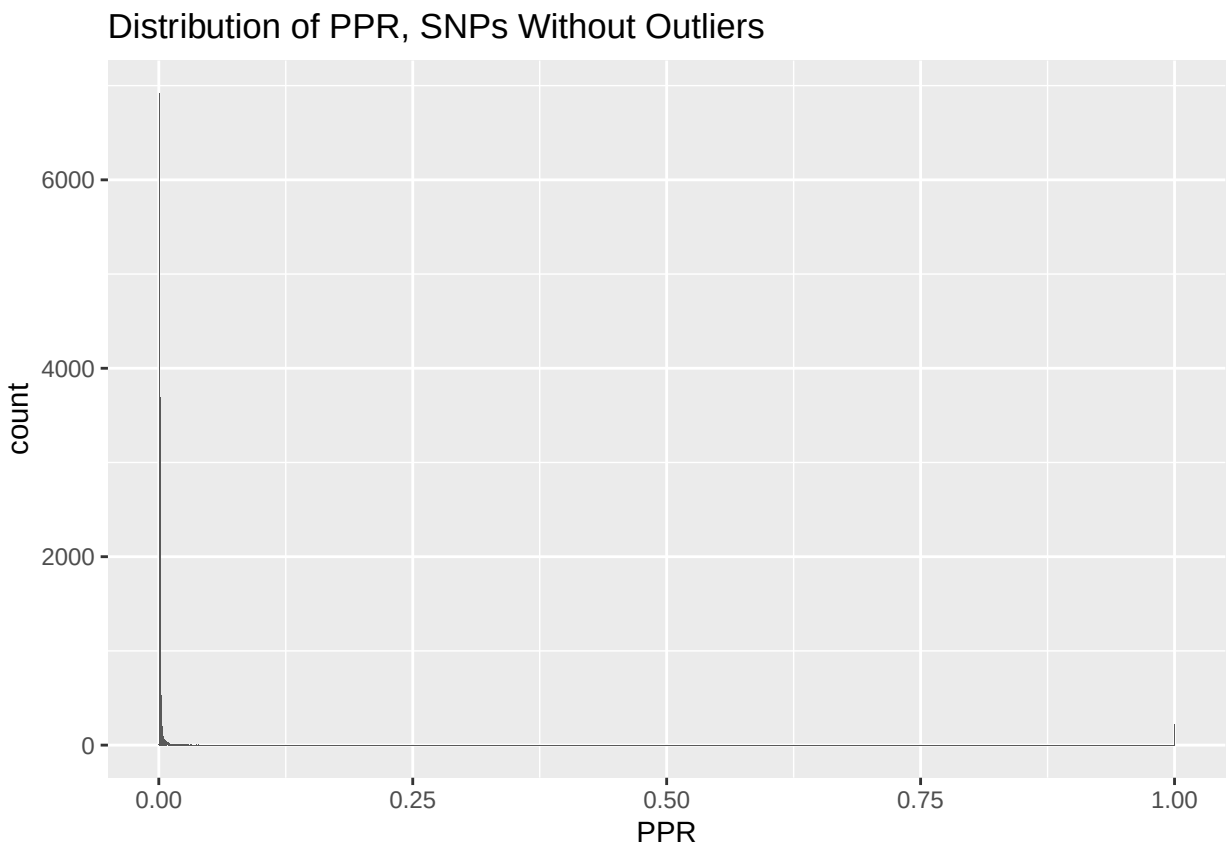


With outliers, there are around 11004 SNPS with PPR values less than or equal to 0.05.

SNPs Without Outliers

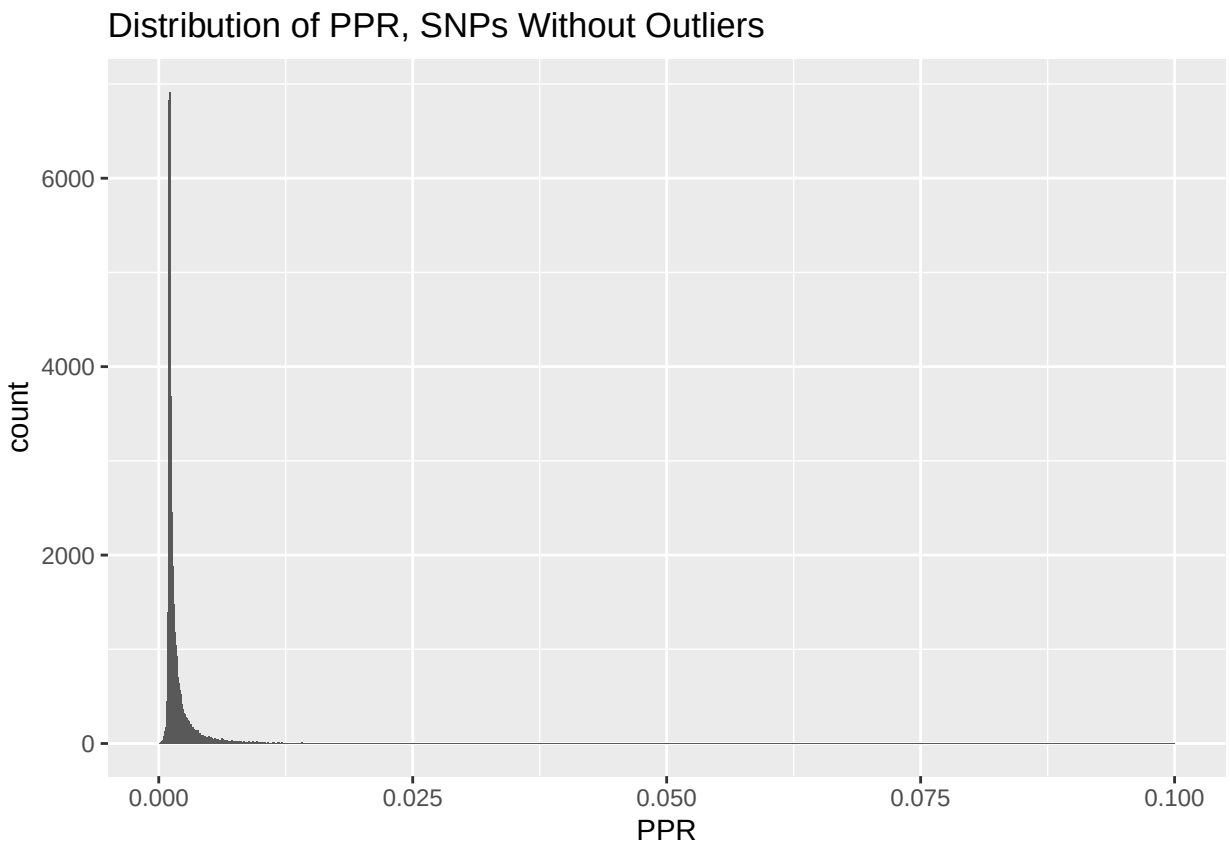
For our SNPs **without outliers**, the PRP distribution is given in the following histogram.

```
ggplot(data = as.data.frame(nonout_ppr), aes(nonout_ppr)) +
  geom_histogram(binwidth = 0.0001) +
  ggtitle("Distribution of PPR, SNPs Without Outliers") +
  xlab("PPR")
```



Like before, if we ignore extreme PPR values, our histogram looks different.

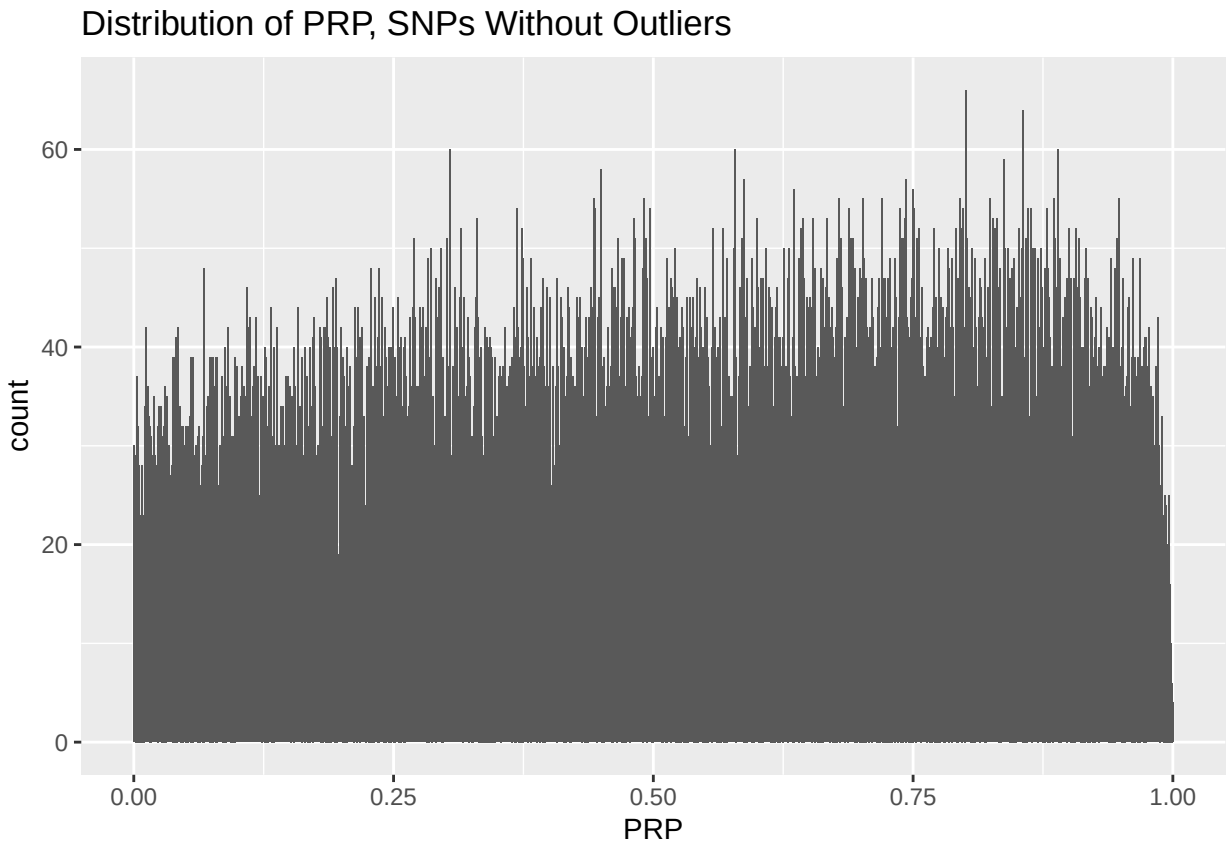
```
ggplot(data = as.data.frame(nonout_ppr), aes(nonout_ppr)) +
  geom_histogram(binwidth = 0.0001) +
  ggtitle("Distribution of PPR, SNPs Without Outliers") +
  xlab("PPR") +
  xlim(0, 0.1)
```



The number of known nonoutlier SNPs with PPR values less than or equal to 0.05 is 38376. The total number of SNPs with PPRs less than or equal to 0.05 using the MAMBA method is 49380.

In contrast, this is what the distribution of our PRP values looks like.

```
ggplot(data = as.data.frame(nonout_prp), aes(nonout_prp)) +
  geom_histogram(binwidth = 0.001) +
  ggtitle("Distribution of PRP, SNPs Without Outliers") +
  xlab("PRP")
```



The number of known nonoutlier SNPs with PPR values less than or equal to 0.05 is 38376. The total number of SNPs with PPRs less than or equal to 0.05 using the MAMBA method is 12554. Recall that here, our nonoutlier study rate is 0.975.

Inverse Variance Weighted

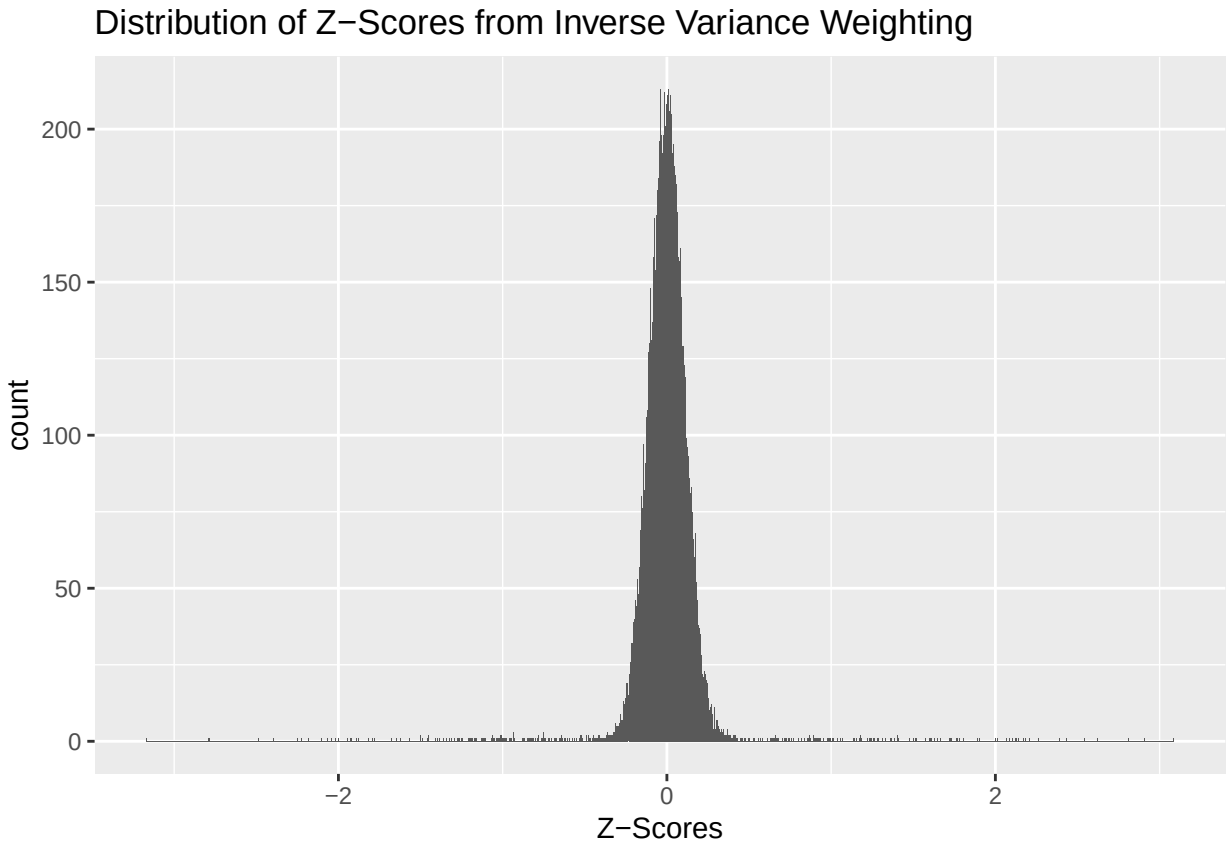
We try an inverse variance weighted method to estimate the beta parameters and get z-score estimates.

Below is a snippet of the data frame containing our inverse varaince weighted estimates and their respective z-scores.

beta	se	zscore
0.001847	0.01439	0.1283
0.002538	0.01444	0.1758
-0.0003636	0.01429	-0.02544
-0.0008766	0.01442	-0.06078
0.003212	0.01448	0.2218

Below, we see the distribution of z-scores from inverse variance weighting.

```
ggplot(ivw_df, aes(x = zscore)) +
  geom_histogram(binwidth = 0.001) +
  ggtitle("Distribution of Z-Scores from Inverse Variance Weighting") +
  xlab("Z-Scores")
```



Giant Consortium Studies

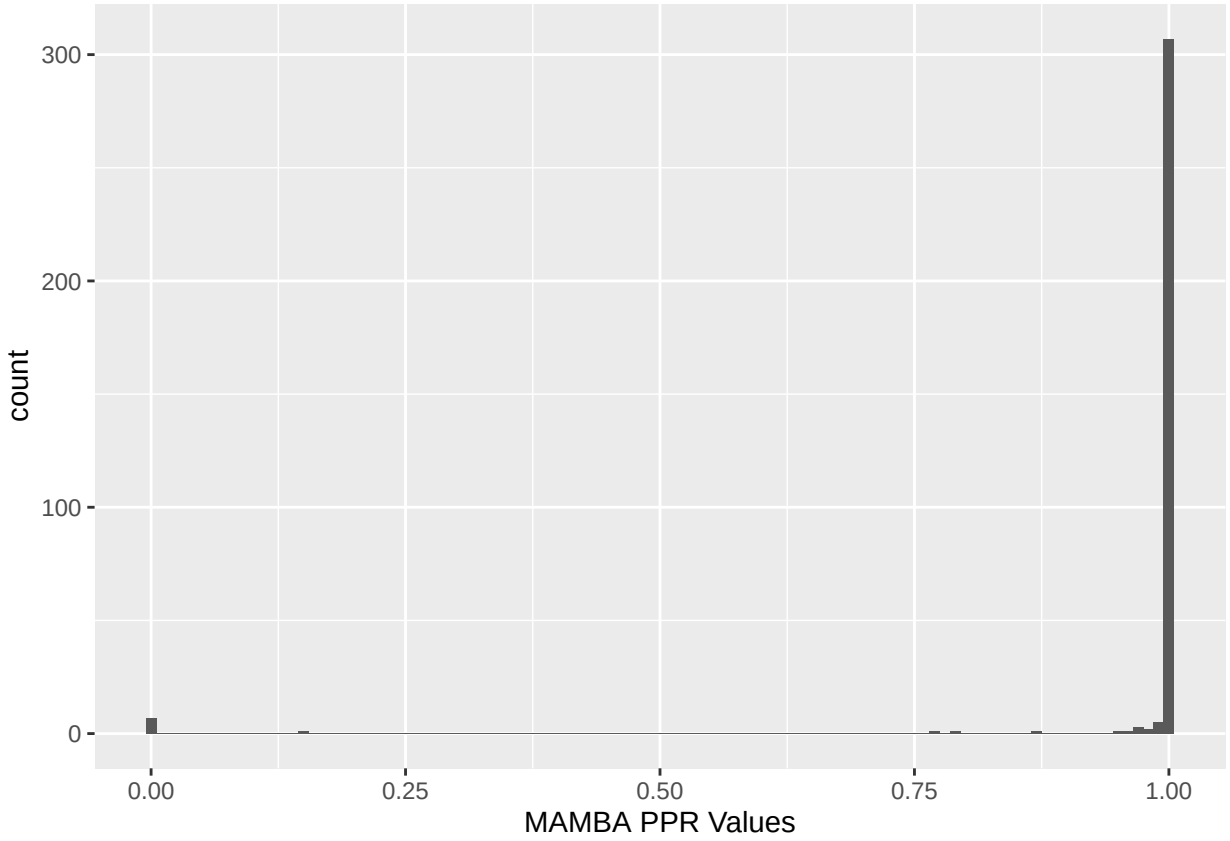
BMI

First we look at the BMI. We'll only look at the SNPs that are significant (where the p-values for "All" data file is less than 0.05/the number of SNPs).

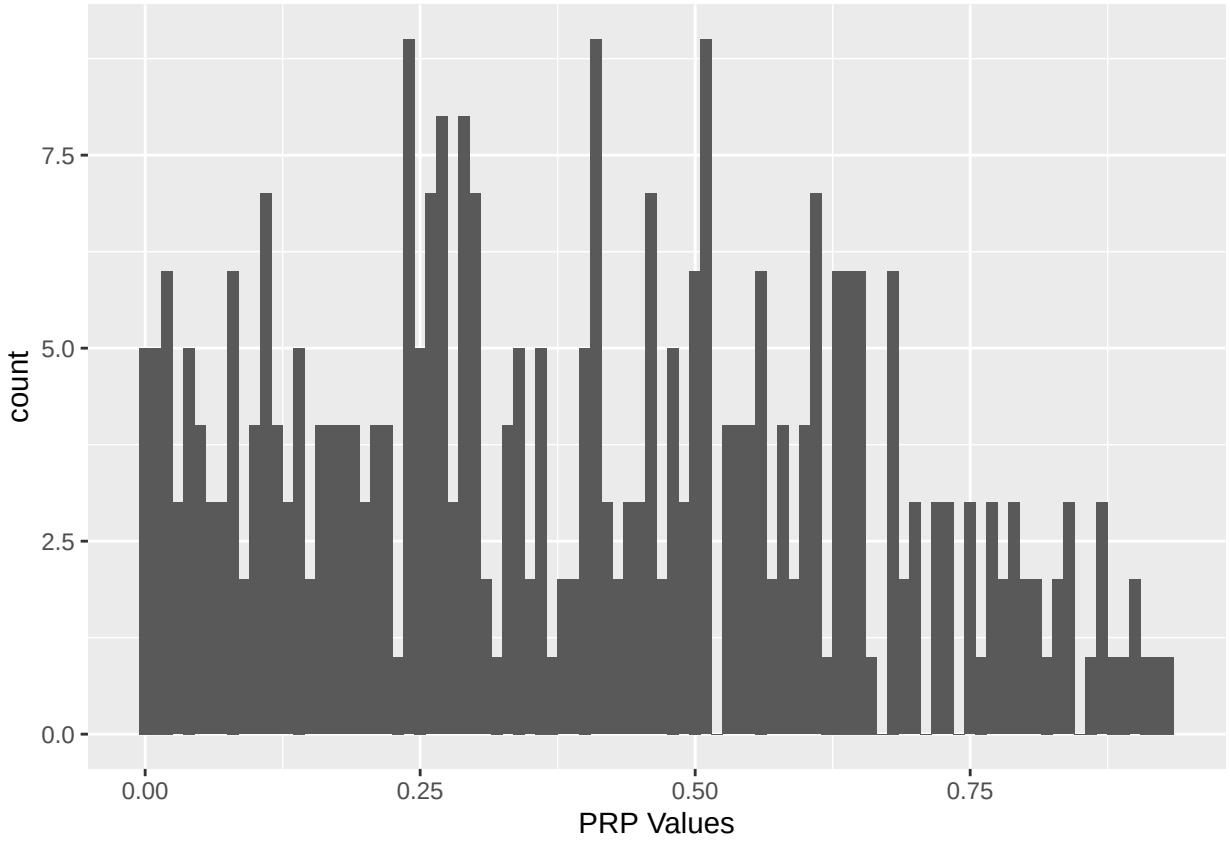
BMI Graphs

Now we get a better look at the BMI through graphical representations.

We take a look at our PPR and PRP values. The histogram of our PPR values for BMI is given below.



In contrast, here is our histogram of PRP values.

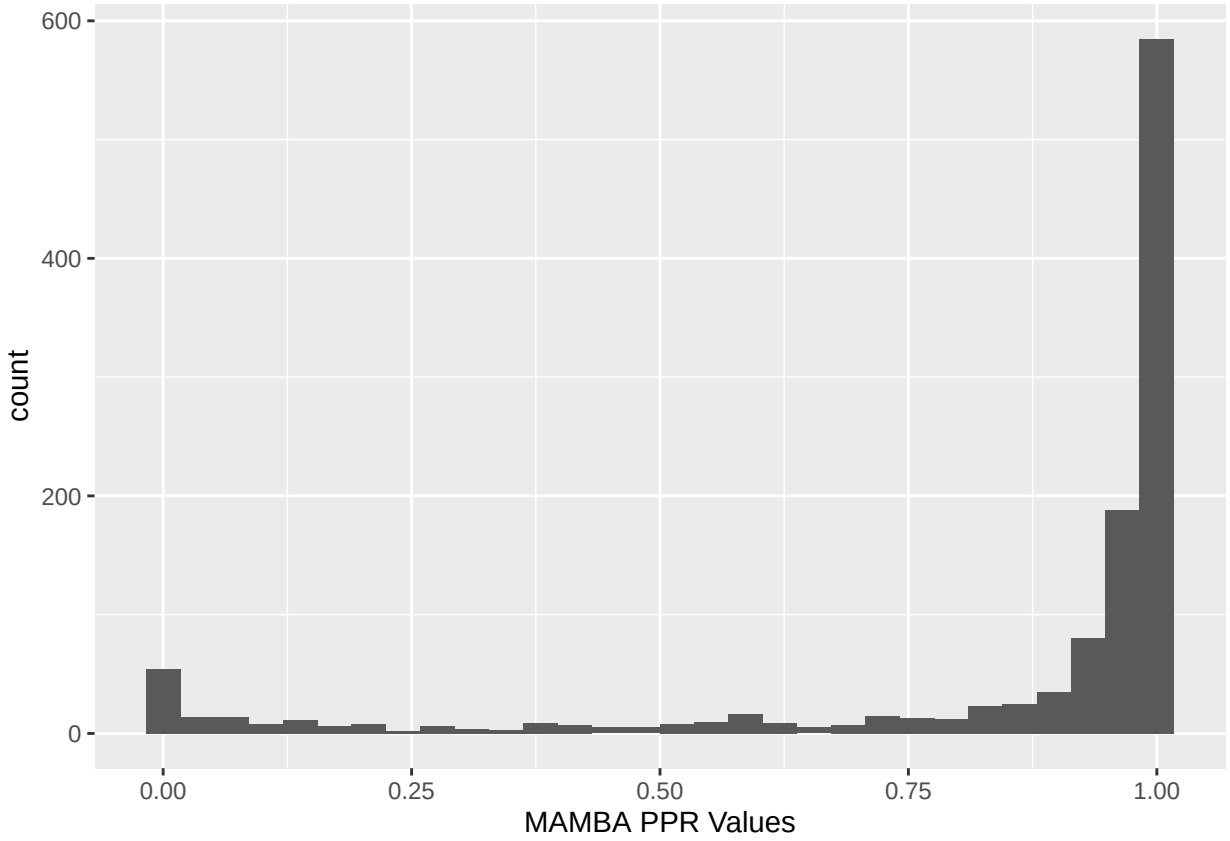

7.5
5.0
2.5
0.0
count
0.00
0.25
0.50
0.75
PRP Values

Height

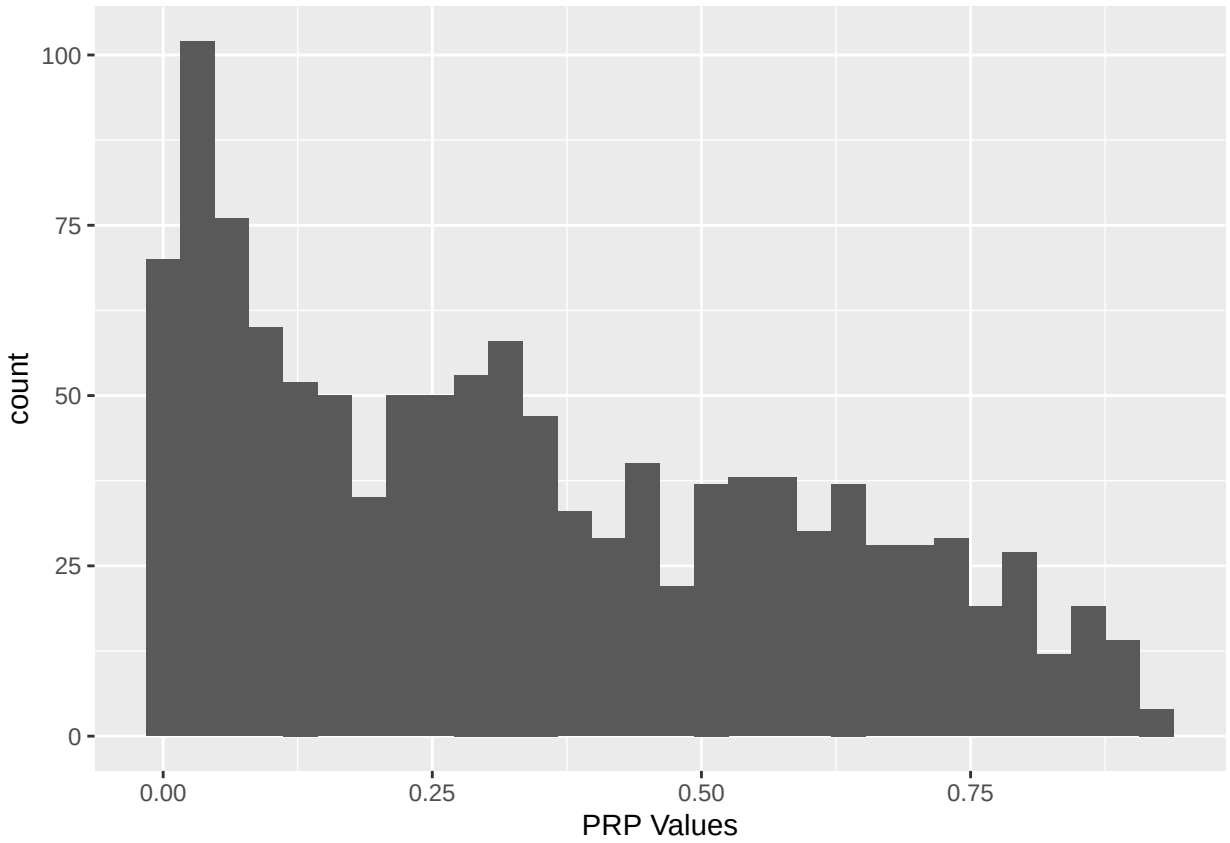
Next we look at height.

Height Graphs

We take a look at our PPR and PRP values. The histogram of our PPR values is given below.



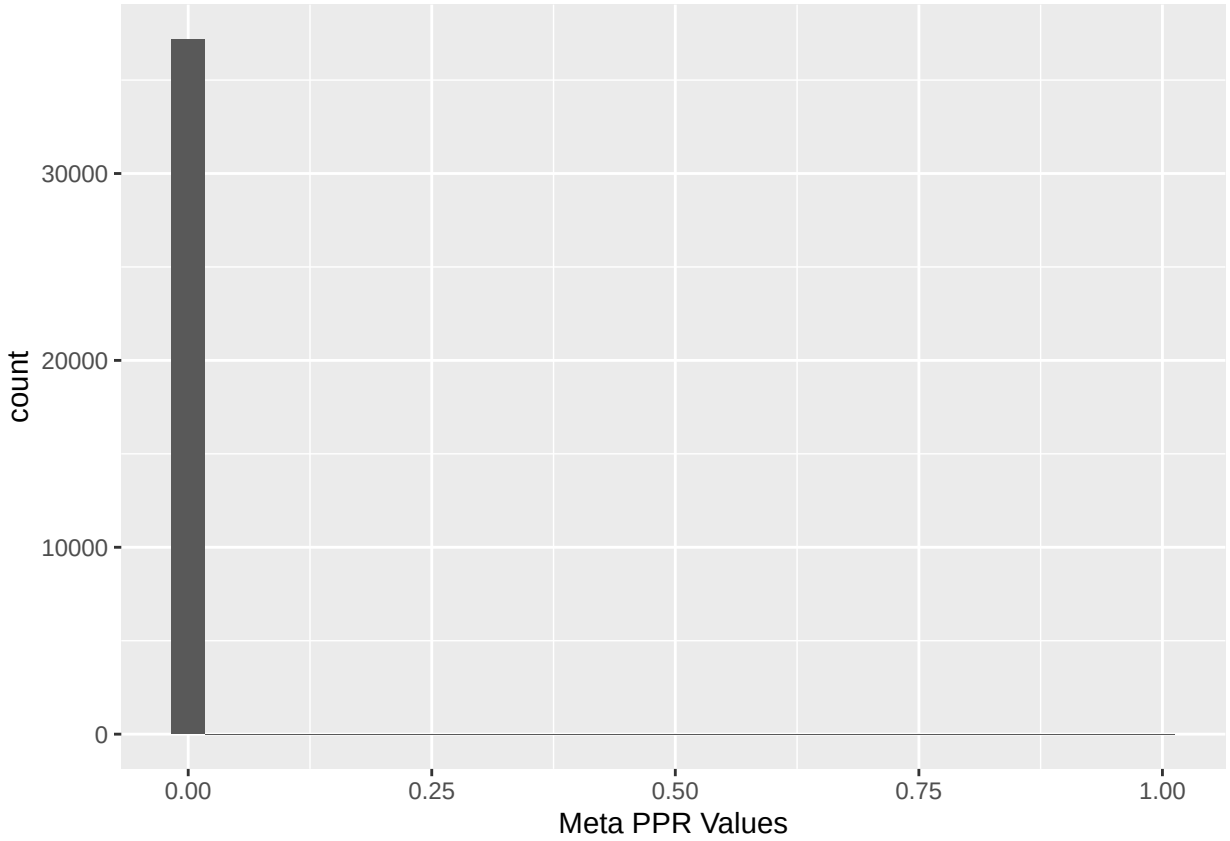
In contrast, we have our histogram for PRP values.



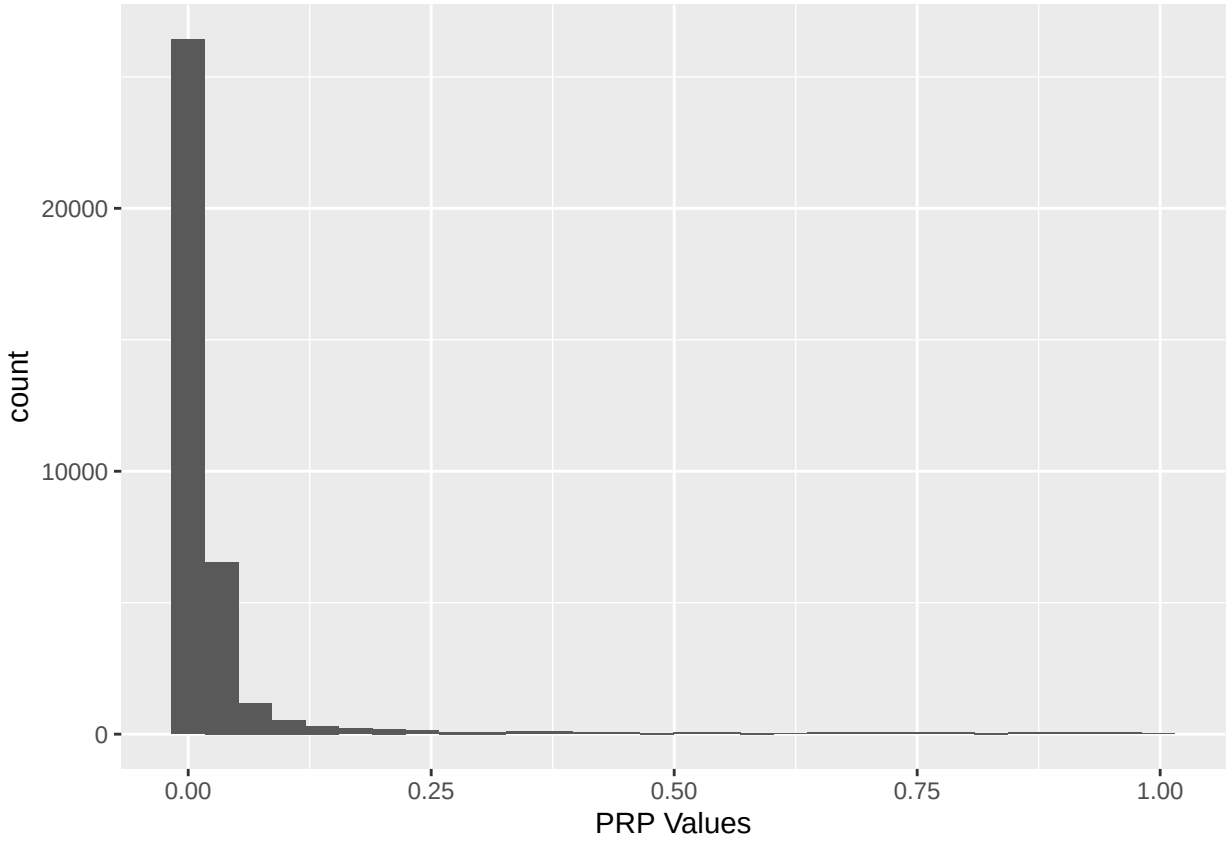
Prior Checking

Meta Graphs

We take a look at our PPR and PRP values. The histogram of our PPR values is given below.



In contrast, we have our histogram for PRP values.


20000
10000
0
count
0.00
0.25
0.50
0.75
1.00
PRP Values