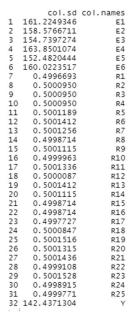
Nicholas Barrett 110355429

## **Summary Statistics**

> summary(dataset							
E1	E2	E3	E4	E5	E6	R1	R2
Min. : 14.51	Min. :-232.8	Min. : 453.7	Min. : 400.4	Min. : 490.4	Min. : 17.02	Min. :0.000	Min. :0.0000
1st Qu.: 455.52	1st Qu.: 207.3	1st Qu.: 828.7	1st Qu.: 758.2	1st Qu.: 965.1	1st Qu.: 378.26	1st Qu.:0.000	1st Qu.:0.0000
Median : 566.37	Median : 316.0	Median : 932.4	Median : 866.1	Median :1063.4	Median : 490.49	Median :0.000	Median :0.0000
Mean : 566.15	Mean : 313.7	Mean : 934.0	Mean : 867.4	Mean :1067.3	Mean : 485.51	Mean :0.478	Mean :0.4922
3rd Qu.: 679.59	3rd Qu.: 415.5	3rd Qu.:1035.3	3rd Qu.: 975.5	3rd Qu.:1171.1	3rd Qu.: 591.73	3rd Qu.:1.000	3rd Qu.:1.0000
Max. :1083.97	Max. : 759.6	Max. :1451.3	Max. :1329.5	Max. :1574.9	Max. :1057.54	Max. :1.000	Max. :1.0000
NA's :20	NA's :30	NA's :20	NA's :20	NA's :30	NA's :20	NA's :30	NA's :30
R3	R4	R5	R6	R7	R8	R9	R10
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :1.0000	Median :1.0000	Median :1.0000	Median :0.0000	Median :0.0000	Median :1.0000	Median :0.0000	Median :1.0000
Mean :0.5078	Mean :0.5078	Mean :0.5058	Mean :0.4966	Mean :0.4948	Mean :0.5168	Mean :0.4936	Mean :0.5125
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000
NA'S :30	NA's :30						
R11	R12	R13		R15	R16	R17	
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :1.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :1.0000	Median :1.0000	Median :1.0000	Median :1.0000
Mean :0.5047	Mean :0.4879	Mean :0.4966	Mean :0.4936	Mean :0.5168	Mean :0.5168	Mean :0.5196	Mean :0.5082
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000
NA's :30	NA's :30					NA's :30	
R19	R20	R21	R22	R23	R24	R25	Υ
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :2418
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:2817
Median :1.0000	Median :0.0000	Median :1.0000	Median :1.0000	Median :0.0000	Median :0.0000	Median :1.0000	Median :2910
Mean :0.5009	Mean :0.4954	Mean :0.5034	Mean :0.5155	Mean :0.4984	Mean :0.4839	Mean :0.5133	Mean :2909
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:3009
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :3320
		NA's :30		NA's :30		NA's :30	NA's :30

**Figure 1 Summary Statistics** 

There were 1641 observations across the 32 variables. The standard deviations for these variables are in the following table. For these statistics, the NA values were omitted.



**Figure 2 Standard Deviations** 

## **Missing Values**

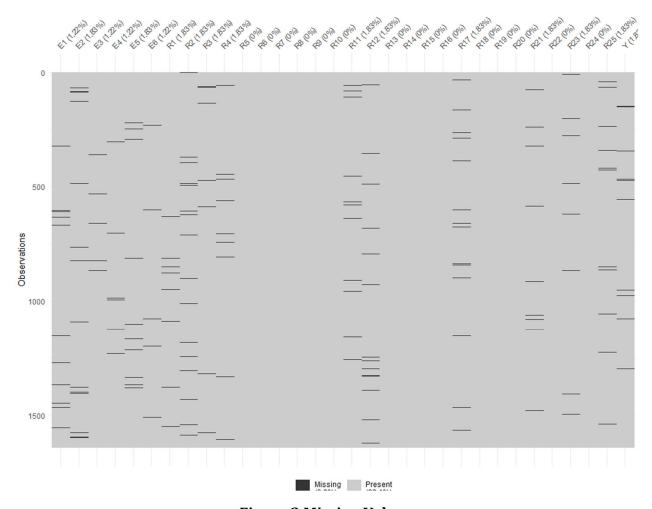
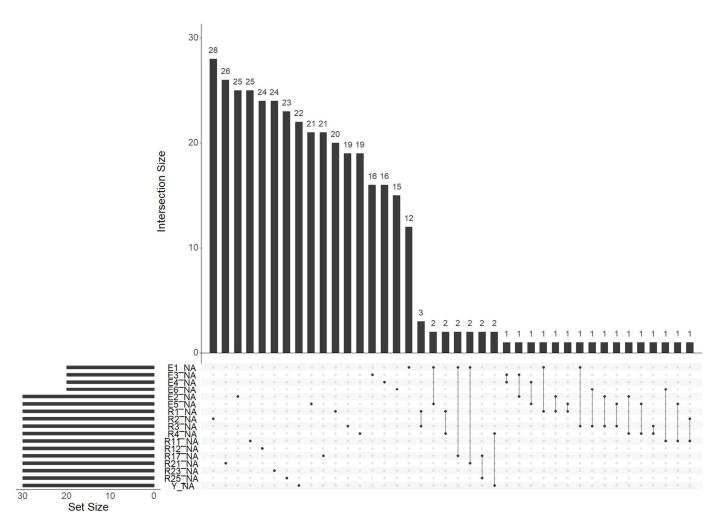


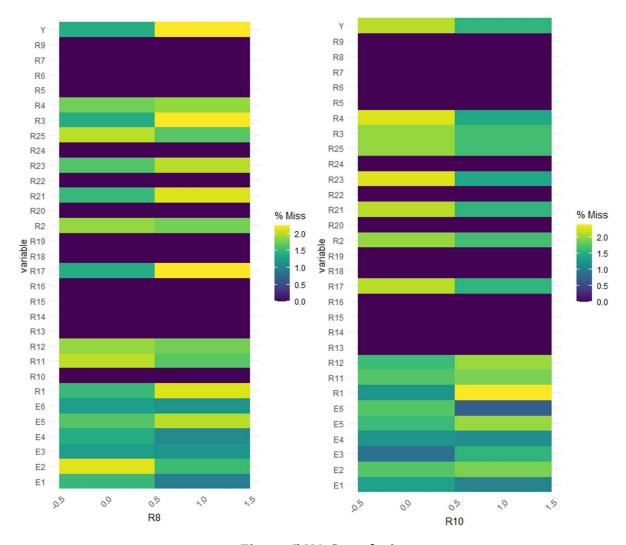
Figure 3 Missing Values

From the summary statistics and this missing value plot we can see several variables are free of missing values, and that there is no obvious pattern in the missing values, we can inspect the intersection of missing values in the next figure.



**Figure 4 Missing Value Intersections** 

Here we can see that the majority of missing values do not intersect in rows with other missing values, and that intersections rarely occur more than once.



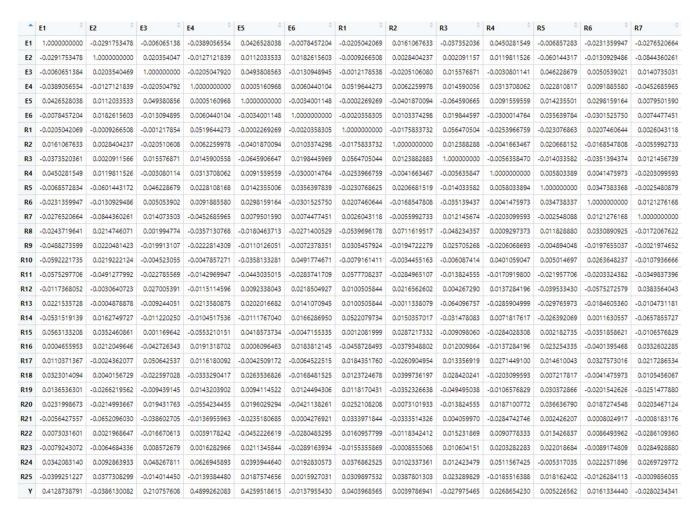
**Figure 5 NA Correlation** 

Some of the genetic indicator factor variables correlate with number of missing variables in unique ways across the rest of the variables. This is an example of R8 and R10's correlation with missing values, these were among the highest correlations for missing values, but the sample sizes are small with the number of NA's either being 30 or 20 for variables. It is hard to tell if there is an overarching pattern or not.

## **Imputation Method**

For imputation, each incomplete variable will be imputed by a separate model created by classification and regression trees (CART) by Gibbs sampling. The implantation of the method being used is the MICE package in R. This will generate multiple imputations for each column based on models of the other columns, which are then evaluated and pooled after analysis for a robust imputation. If necessary, a separate univariate imputation model can be used for each column/variable, such as a random forest-based model or logistic regression for the categorical variables. This will depend on subsequent analysis.

## **Correlation Matrix**



**Figure 6 Correlation Matrix 1** 

<b>R7</b> <sup>‡</sup>	R8 <sup>‡</sup>	R9 <sup>‡</sup>	R10 <sup>‡</sup>	R11 <sup>‡</sup>	R12 <sup>‡</sup>	R13 <sup>‡</sup>	R14 <sup>‡</sup>	R15 <sup>‡</sup>	R16 <sup>‡</sup>	R17 <sup>‡</sup>	R18 <sup>‡</sup>
-0.0276520664	-0.0243719641	-0.048827360	-0.059222174	-0.0575297706	-0.011736805	0.0221535728	-0.0531519139	0.056313321	0.0004655953	0.011037137	0.032301409
-0.0844360261	0.0214746071	0.022048142	0.021922212	-0.0491277992	-0.003064072	-0.0004878878	0.0162749727	0.035246086	0.0212049646	-0.002436208	0.004015673
0.0140735031	0.0019947742	-0.019913107	-0.004523055	-0.0227855685	0.027005391	-0.0092440510	-0.0112202503	0.001169642	-0.0427263431	0.050642537	-0.022597028
-0.0452685965	-0.0357130768	-0.022281431	-0.004785727	-0.0142969947	-0.011511460	0.0213580875	-0.0104517536	-0.055321015	0.0191318702	0.011618009	-0.033329042
0.0079501590	-0.0180463713	-0.011012605	-0.035813328	-0.0443035015	0.009233804	0.0202016682	-0.0111767040	0.041857373	0.0006096463	-0.004250917	0.026353683
0.0074477451	-0.0271400529	-0.007237835	0.049177467	-0.0283741709	0.021850493	0.0141070945	0.0166286950	-0.004715533	0.0183812145	-0.006452251	-0.016848152
0.0026043118	-0.0539696178	0.030545792	-0.007916141	0.0577708237	0.010050584	0.0100505844	0.0522079734	0.001208200	-0.0458728493	0.018435176	0.012372468
-0.0055992733	0.0711619517	-0.019472228	-0.003445516	-0.0284965107	0.021656260	-0.0011338079	0.0150357017	0.028721733	-0.0379348802	-0.026090495	0.039973620
0.0121456739	-0.0482343566	0.025705268	-0.006087414	-0.0138245553	0.004267290	-0.0640967575	-0.0314780827	-0.009098060	0.0120098644	0.013356919	0.028420241
-0.0203099593	0.0009297373	-0.020606869	0.040105905	-0.0170919800	0.013728420	-0.0285904999	0.0071817617	-0.028402831	-0.0137284196	0.027144910	-0.020309959
-0.0025480879	0.0118288800	-0.004894048	0.005014697	-0.0219577059	-0.039533430	-0.0297659731	-0.0263920686	0.002182735	0.0232543354	0.014610043	0.007217817
0.0121276168	0.0330890925	-0.019765504	0.026364824	-0.0203324382	-0.057527258	-0.0184605360	0.0011630557	-0.035185862	-0.0401395468	0.032757302	-0.004147597
1.0000000000	-0.0172067622	-0.002197465	-0.010793667	-0.0349837396	0.038356404	-0.0104731181	-0.0657855727	-0.010657683	0.0332602285	0.021728653	0.010545607
-0.0172067622	1.0000000000	-0.009711520	0.003326451	0.0187348361	0.046777723	-0.0215966471	-0.0279800326	-0.002723208	0.0118288800	-0.016424225	-0.036739192
-0.0021974652	-0.0097115204	1.000000000	-0.035092281	0.0203255676	-0.017913911	0.0211854469	0.0046136844	0.003040643	0.0048807919	-0.015117301	0.014091345
-0.0107936666	0.0033264511	-0.035092281	1.000000000	0.0285273691	0.043846797	0.0112724675	-0.0015189391	-0.022492788	0.0310741614	0.034978024	-0.040105905
-0.0349837396	0.0187348361	0.020325568	0.028527369	1.0000000000	-0.010594792	0.0187024562	-0.0008344592	-0.007312474	0.0105947918	0.025288718	0.056148770
0.0383564043	0.0467777231	-0.017913911	0.043846797	-0.0105947918	1.000000000	0.0265101542	0.0035980754	-0.002182735	-0.0265101542	-0.017868798	0.012313992
-0.0104731181	-0.0215966471	0.021185447	0.011272467	0.0187024562	0.026510154	1.0000000000	-0.0582713347	0.007583945	-0.0004636032	-0.027645064	-0.007217817
-0.0657855727	-0.0279800326	0.004613684	-0.001518939	-0.0008344592	0.003598075	-0.0582713347	1.0000000000	0.040235367	-0.0329046381	-0.017569035	0.009097075
-0.0106576829	-0.0027232077	0.003040643	-0.022492788	-0.0073124739	-0.002182735	0.0075839453	0.0402353674	1.000000000	-0.0238617461	-0.016120145	-0.039953068
0.0332602285	0.0118288800	0.004880792	0.031074161	0.0105947918	-0.026510154	-0.0004636032	-0.0329046381	-0.023861746	1.0000000000	-0.034271289	0.003962515
0.0217286534	-0.0164242252	-0.015117301	0.034978024	0.0252887178	-0.017868798	-0.0276450642	-0.0175690348	-0.016120145	-0.0342712886	1.000000000	-0.023886672
0.0105456067	-0.0367391921	0.014091345	-0.040105905	0.0561487696	0.012313992	-0.0072178166	0.0090970747	-0.039953068	0.0039625151	-0.023886672	1.000000000
-0.0251477880	0.0515581625	0.006733420	-0.010078953	0.0040574828	0.005438295	-0.0140950656	0.0020929703	-0.013643660	-0.0086938555	0.025895635	-0.057698216
0.0203467124	0.0415223089	0.023583278	0.028527369	-0.0024416726	0.018702456	0.0284682055	-0.0236244681	0.054532357	0.0138500415	-0.033358626	-0.038238472
-0.0008183176	-0.0089687763	0.031797797	0.039866766	0.0447525224	-0.028468205	-0.0089367069	-0.0121884030	-0.021982446	0.0349787050	0.043133184	-0.030110910
-0.0286109360	0.0085072899	-0.040597303	-0.008134845	-0.0056791845	0.058205922	-0.0004026991	-0.0002596721	-0.059721895	-0.0061093699	-0.005093966	0.003944235
0.0284928880	0.0334171241	-0.072589612	-0.044692555	-0.0073250373	-0.002487134	0.0365759667	0.0040080640	0.030142516	0.0057423923	0.010702466	0.025238147
0.0269729772	0.0321602103	-0.022872314	0.006446159	-0.0122241296	0.037876259	0.0508999486	-0.0208657914	-0.020066438	-0.0150848023	0.016424225	-0.008836478
-0.0009856055	0.0067150555	0.039595537	-0.078489129	0.0220035578	-0.008392228	0.0046346720	0.0048314557	-0.009371945	0.0051355027	-0.023401139	-0.010754228
-0.0280234341	-0.0110401946	-0.046392020	-0.035517125	-0.0375238023	0.020574221	0.0233505340	-0.0336172394	-0.007838996	-0.0220443279	0.030335120	-0.018107955

Figure 7 Correlation Matrix 2

R19 <sup>‡</sup>	R20 <sup>‡</sup>	R21 <sup>‡</sup>	R22 <sup>‡</sup>	R23 <sup>‡</sup>	R24 <sup>‡</sup>	R25 <sup>‡</sup>	<b>Y</b>
0.013653630	0.023199867	-0.0056427557	0.0073031601	-0.0079243072	0.034208314	-0.0399251227	0.412873879
-0.026621956	-0.021499367	-0.0652096030	0.0021968647	-0.0064684336	0.009286393	0.0377308299	-0.038613008
-0.009439145	0.019431763	-0.0386027046	-0.0166706126	0.0085726788	0.048267811	-0.0140144500	0.210757608
0.014320390	-0.055423445	-0.0136955963	0.0039178242	0.0016282966	0.062694589	-0.0139384480	0.489926208
0.009411452	0.019602929	-0.0235180685	-0.0452226619	0.0211345844	0.039394464	0.0187574656	0.425951862
0.012449431	-0.042113826	0.0004276921	-0.0280483295	-0.0289163934	0.019283057	0.0015927031	-0.013795543
0.011817043	0.025210821	0.0333971844	0.0160957799	-0.0155355869	0.037686252	0.0309897532	0.040396856
-0.035232664	0.007310193	-0.0333514326	-0.0118342412	-0.0008555068	0.010233736	0.0387801303	0.003978694
-0.049495038	-0.013824555	0.0040599700	0.0152318690	0.0106041506	0.012423479	0.0232898290	-0.027975465
-0.010657683	0.018710077	-0.0284742746	0.0090778333	0.0203282283	0.051156742	-0.0185516388	0.026865423
0.030372866	0.036636790	0.0024262073	0.0134268373	0.0220186843	-0.005317035	0.0181624024	0.005226562
-0.020154263	0.018727455	0.0008024917	0.0086493962	-0.0089174809	0.022257190	-0.0126284113	0.016133444
-0.025147788	0.020346712	-0.0008183176	-0.0286109360	0.0284928880	0.026972977	-0.0009856055	-0.028023434
0.051558163	0.041522309	-0.0089687763	0.0085072899	0.0334171241	0.032160210	0.0067150555	-0.011040195
0.006733420	0.023583278	0.0317977972	-0.0405973032	-0.0725896118	-0.022872314	0.0395955366	-0.046392020
-0.010078953	0.028527369	0.0398667664	-0.0081348446	-0.0446925549	0.006446159	-0.0784891292	-0.035517125
0.004057483	-0.002441673	0.0447525224	-0.0056791845	-0.0073250373	-0.012224130	0.0220035578	-0.037523802
0.005438295	0.018702456	-0.0284682055	0.0582059225	-0.0024871340	0.037876259	-0.0083922276	0.020574221
-0.014095066	0.028468205	-0.0089367069	-0.0004026991	0.0365759667	0.050899949	0.0046346720	0.023350534
0.002092970	-0.023624468	-0.0121884030	-0.0002596721	0.0040080640	-0.020865791	0.0048314557	-0.033617239
-0.013643660	0.054532357	-0.0219824459	-0.0597218954	0.0301425156	-0.020066438	-0.0093719452	-0.007838996
-0.008693856	0.013850042	0.0349787050	-0.0061093699	0.0057423923	-0.015084802	0.0051355027	-0.022044328
0.025895635	-0.033358626	0.0431331837	-0.0050939657	0.0107024661	0.016424225	-0.0234011392	0.030335120
-0.057698216	-0.038238472	-0.0301109098	0.0039442351	0.0252381469	-0.008836478	-0.0107542276	-0.018107955
1.000000000	-0.028492428	0.0284924281	-0.0216724994	-0.0138675171	-0.015745862	-0.0166797837	-0.015917939
-0.028492428	1.000000000	-0.0203410927	-0.0284674422	-0.0268531735	-0.021990189	0.0024666240	-0.002910912
0.028492428	-0.020341093	1.0000000000	-0.0268754693	0.0333625523	-0.013818696	-0.0252597134	-0.032557124
-0.021672499	-0.028467442	-0.0268754693	1.0000000000	-0.0398259167	0.063127747	-0.0112469129	-0.001925618
-0.013867517	-0.026853174	0.0333625523	-0.0398259167	1.0000000000	0.008902580	-0.0414425723	-0.021286244
-0.015745862	-0.021990189	-0.0138186965	0.0631277472	0.0089025804	1.000000000	-0.0164855409	0.068907284
-0.016679784	0.002466624	-0.0252597134	-0.0112469129	-0.0414425723	-0.016485541	1.0000000000	-0.048812037
-0.015917939	-0.002910912	-0.0325571243	-0.0019256183	-0.0212862441	0.068907284	-0.0488120368	1.000000000

Figure 8 Correlation Matrix 3