statgen_sibs

Jayden Chrzanowski

2022-06-29

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(arsenal)
```

```
## Warning: package 'arsenal' was built under R version 4.1.3
```

```
# reading phenotype data
phenotype <- read.csv("T2D-GENES_P1_Hispanic_phenotypes.txt", sep="")
head(phenotype)
```

```
##           FID         IID DBP SBP BPMED SMOKE EXAM_YEAR AGE SEX
## 1 SAMPLE_0001 SAMPLE_0001  51 103     0    NA      2000  57   1
## 2 SAMPLE_0002 SAMPLE_0002 109 184     1    NA      1999  60   1
## 3 SAMPLE_0003 SAMPLE_0003  92 182     1    NA      1999  50   2
## 4 SAMPLE_0004 SAMPLE_0004  78 141     1    NA      2000  59   2
## 5 SAMPLE_0005 SAMPLE_0005  89 150     1    NA      1999  41   2
## 6 SAMPLE_0006 SAMPLE_0006  57 100     0    NA      1997  75   2
```

```
# reading genetic data
pcs <- read.csv("all_chr_eigensoft_PCs.csv")
head(pcs)
```

```
##           ID     PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8
## 1 SAMPLE_0001 -0.0613 -0.0033 -0.0087  0.0194  0.0182 -0.0194 -0.0244  0.0149
## 2 SAMPLE_0002 -0.0144 -0.0126  0.0248 -0.0131 -0.0207 -0.0316  0.0166  0.0361
```

```
## 3 SAMPLE_0003 -0.0217 -0.0124  0.0176 -0.0160 -0.0186 -0.0101  0.0191  0.0103
## 4 SAMPLE_0004 -0.0130 -0.0079 -0.0013  0.0284  0.0304 -0.0025  0.0061  0.0117
## 5 SAMPLE_0005 -0.0136 -0.0163 -0.0433 -0.0469 -0.0417  0.0141  0.0006 -0.0175
## 6 SAMPLE_0006 -0.0185 -0.0139  0.0091  0.0035  0.0029 -0.0132 -0.0029  0.0035
##       PC9    PC10
## 1  0.0180 -0.0323
## 2 -0.0100  0.0269
## 3 -0.0093 -0.0054
## 4 -0.0001 -0.0201
## 5  0.0310  0.0236
## 6 -0.0091 -0.0102
```

```
phenotype <- subset(phenotype, select = -SMOKE) #removing smoke variable - all NAs
phenotype$HTN <- ifelse(phenotype$SBP>130, 1, ifelse(is.null(phenotype$BPMED), 1, ifelse(phenotype$BPMED
phenotype <- rename(phenotype, "ID" = "IID") # renaming IID in phenotype data to ID to merge with genet
phenotype <- merge(phenotype, pcs, by="ID") # merging genetic data by id var
```

```
head(phenotype)
```

```
##            ID         FID DBP SBP BPMED EXAM_YEAR AGE SEX HTN     PC1     PC2
## 1 SAMPLE_0001 SAMPLE_0001  51 103     0      2000  57   1   0 -0.0613 -0.0033
## 2 SAMPLE_0002 SAMPLE_0002 109 184     1      1999  60   1   1 -0.0144 -0.0126
## 3 SAMPLE_0003 SAMPLE_0003  92 182     1      1999  50   2   1 -0.0217 -0.0124
## 4 SAMPLE_0004 SAMPLE_0004  78 141     1      2000  59   2   1 -0.0130 -0.0079
## 5 SAMPLE_0005 SAMPLE_0005  89 150     1      1999  41   2   1 -0.0136 -0.0163
## 6 SAMPLE_0006 SAMPLE_0006  57 100     0      1997  75   2   0 -0.0185 -0.0139
##       PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10
## 1 -0.0087  0.0194  0.0182 -0.0194 -0.0244  0.0149  0.0180 -0.0323
## 2  0.0248 -0.0131 -0.0207 -0.0316  0.0166  0.0361 -0.0100  0.0269
## 3  0.0176 -0.0160 -0.0186 -0.0101  0.0191  0.0103 -0.0093 -0.0054
## 4 -0.0013  0.0284  0.0304 -0.0025  0.0061  0.0117 -0.0001 -0.0201
## 5 -0.0433 -0.0469 -0.0417  0.0141  0.0006 -0.0175  0.0310  0.0236
## 6  0.0091  0.0035  0.0029 -0.0132 -0.0029  0.0035 -0.0091 -0.0102
```

```
colnames(phenotype)
```

```
##  [1] "ID"        "FID"       "DBP"       "SBP"       "BPMED"     "EXAM_YEAR"
##  [7] "AGE"       "SEX"       "HTN"       "PC1"       "PC2"       "PC3"
## [13] "PC4"       "PC5"       "PC6"       "PC7"       "PC8"       "PC9"
## [19] "PC10"
```

```
# Controls for summary stat tables
my_controls <- tableby.control(
  test = T,
  total = T,
  numeric.test="kwt", cat.test="chisq", #Specify Kruskal-Wallis test for numeric vars and chi square fo
  numeric.stats = c("meansd", "medianq1q3", "range", "Nmiss2"), #Display mean, sd, median, q1 and q3, r
  cat.stats = c("countpct", "Nmiss2"),
  stats.labels=list(
    meansd = "Mean (SD)",
    medianq1q3 = "Median (Q1, Q3)",
    range = "Min - Max",
```

```
    Nmiss2 = "Missing"
  )
)
# Creating labels for variables in table
my_labels <- list(
  DBP = "Diastolic Blood Pressure",
  SBP = "Systolic Blood Pressure",
  AGE = "Age",
  SEX = "Sex",
  BPMED = "Blood Pressure Medication Status",
  HTN = "Hypertension Status"
)

# Creating a list of the PC variables
pc_list <- paste("PC", 1:10, sep="")
htn_fmla <- as.formula(paste("HTN ~ + DBP + SBP + AGE +", paste(pc_list, collapse= "+")))
sex_fmla <- as.formula(paste("SEX ~ + DBP + SBP + AGE +", paste(pc_list, collapse= "+")))
bpmed_fmla <- as.formula(paste("BPMED ~ + DBP + SBP + AGE +", paste(pc_list, collapse= "+")))

htn_table <- tableby(htn_fmla, data = phenotype, control=my_controls)
summary(htn_table, labelTranslations = my_labels, title = "Summary Statistics for Phenotypes By Hyperte
```

Table 1: Summary Statistics for Phenotypes By Hypertension

	0 (N=1212)	1 (N=639)	Total (N=1851)	p value
Diastolic Blood Pressure				< 0.001
Mean (SD)	70.144 (8.953)	79.776 (11.802)	73.469 (11.022)	
Median (Q1, Q3)	70.500 (64.000, 76.000)	79.000 (73.000, 86.000)	73.000 (66.000, 80.000)	
Min - Max	32.000 - 98.000	37.000 - 123.000	32.000 - 123.000	
Missing	0	0	0	
Systolic Blood Pressure				< 0.001
Mean (SD)	112.889 (10.623)	148.099 (16.249)	125.044 (21.103)	
Median (Q1, Q3)	114.000 (105.000, 121.000)	143.000 (136.000, 154.500)	122.000 (110.000, 137.000)	
Min - Max	66.000 - 130.000	130.500 - 213.000	66.000 - 213.000	
Missing	0	0	0	
Age				< 0.001
Mean (SD)	45.418 (13.395)	57.025 (12.589)	49.425 (14.233)	
Median (Q1, Q3)	43.000 (35.000, 53.105)	56.590 (49.000, 66.000)	48.000 (38.000, 59.789)	
Min - Max	20.000 - 92.000	19.000 - 91.970	19.000 - 92.000	
Missing	0	0	0	
PC1				< 0.001
Mean (SD)	0.002 (0.022)	-0.002 (0.023)	0.001 (0.023)	
Median (Q1, Q3)	0.005 (-0.015, 0.021)	-0.002 (-0.021, 0.018)	0.003 (-0.017, 0.020)	
Min - Max	-0.070 - 0.046	-0.053 - 0.044	-0.070 - 0.046	
Missing	0	0	0	

	0 (N=1212)	1 (N=639)	Total (N=1851)	p value
PC2				< 0.001
Mean (SD)	0.002 (0.024)	-0.002 (0.021)	0.001 (0.023)	
Median (Q1, Q3)	-0.003 (-0.016, 0.014)	-0.008 (-0.016, 0.007)	-0.005 (-0.016, 0.012)	
Min - Max	-0.035 - 0.109	-0.034 - 0.097	-0.035 - 0.109	
Missing	0	0	0	
PC3				0.090
Mean (SD)	-0.001 (0.023)	0.001 (0.023)	-0.000 (0.023)	
Median (Q1, Q3)	0.001 (-0.016, 0.015)	0.002 (-0.014, 0.016)	0.001 (-0.015, 0.016)	
Min - Max	-0.073 - 0.065	-0.073 - 0.069	-0.073 - 0.069	
Missing	0	0	0	
PC4				0.157
Mean (SD)	0.001 (0.023)	-0.001 (0.023)	0.000 (0.023)	
Median (Q1, Q3)	0.002 (-0.014, 0.016)	-0.001 (-0.016, 0.015)	0.000 (-0.015, 0.016)	
Min - Max	-0.082 - 0.066	-0.069 - 0.063	-0.082 - 0.066	
Missing	0	0	0	
PC5				0.467
Mean (SD)	0.001 (0.024)	-0.001 (0.021)	0.000 (0.023)	
Median (Q1, Q3)	-0.000 (-0.014, 0.015)	0.001 (-0.014, 0.014)	0.000 (-0.014, 0.014)	
Min - Max	-0.080 - 0.098	-0.066 - 0.069	-0.080 - 0.098	
Missing	0	0	0	
PC6				0.714
Mean (SD)	-0.001 (0.024)	0.001 (0.020)	-0.000 (0.023)	
Median (Q1, Q3)	0.002 (-0.012, 0.015)	0.002 (-0.010, 0.013)	0.002 (-0.011, 0.014)	
Min - Max	-0.122 - 0.057	-0.112 - 0.051	-0.122 - 0.057	
Missing	0	0	0	
PC7				0.551
Mean (SD)	0.000 (0.022)	-0.001 (0.024)	0.000 (0.023)	
Median (Q1, Q3)	0.004 (-0.008, 0.014)	0.004 (-0.009, 0.014)	0.004 (-0.009, 0.014)	
Min - Max	-0.115 - 0.052	-0.121 - 0.053	-0.121 - 0.053	
Missing	0	0	0	
PC8				0.823
Mean (SD)	0.000 (0.023)	-0.000 (0.023)	-0.000 (0.023)	
Median (Q1, Q3)	0.001 (-0.015, 0.016)	0.000 (-0.016, 0.015)	0.000 (-0.015, 0.016)	
Min - Max	-0.072 - 0.076	-0.071 - 0.075	-0.072 - 0.076	
Missing	0	0	0	
PC9				0.138
Mean (SD)	0.000 (0.023)	-0.001 (0.023)	0.000 (0.023)	
Median (Q1, Q3)	0.000 (-0.014, 0.015)	-0.002 (-0.015, 0.012)	-0.001 (-0.014, 0.014)	
Min - Max	-0.072 - 0.106	-0.071 - 0.075	-0.072 - 0.106	
Missing	0	0	0	
PC10				0.729
Mean (SD)	-0.000 (0.022)	0.000 (0.023)	-0.000 (0.023)	
Median (Q1, Q3)	-0.000 (-0.016, 0.015)	0.000 (-0.016, 0.016)	0.000 (-0.016, 0.016)	
Min - Max	-0.077 - 0.077	-0.075 - 0.072	-0.077 - 0.077	
Missing	0	0	0	

```
sex_table <- tableby(sex_fmla, data = phenotype, control=my_controls)
summary(sex_table, labelTranslations = my_labels, title = "Summary Statistics for Phenotype Data By Sex
```

Table 2: Summary Statistics for Phenotype Data By Sex

	1 (N=702)	2 (N=1241)	Total (N=1943)	p value
Diastolic Blood Pressure				< 0.001
Mean (SD)	76.619 (11.939)	71.720 (10.067)	73.469 (11.022)	
Median (Q1, Q3)	76.000 (69.000, 83.000)	71.000 (65.000, 78.000)	73.000 (66.000, 80.000)	
Min - Max	42.000 - 123.000	32.000 - 115.000	32.000 - 123.000	
Missing	41	51	92	
Systolic Blood Pressure				< 0.001
Mean (SD)	127.914 (19.799)	123.450 (21.637)	125.044 (21.103)	
Median (Q1, Q3)	125.000 (114.000, 139.000)	120.000 (107.000, 136.000)	122.000 (110.000, 137.000)	
Min - Max	66.000 - 203.000	79.000 - 213.000	66.000 - 213.000	
Missing	41	51	92	
Age				< 0.001
Mean (SD)	50.898 (14.243)	48.700 (14.188)	49.485 (14.243)	
Median (Q1, Q3)	50.110 (40.000, 61.000)	47.000 (38.000, 59.000)	48.000 (38.000, 60.000)	
Min - Max	19.000 - 87.000	22.190 - 92.000	19.000 - 92.000	
Missing	37	44	81	
PC1				0.829
Mean (SD)	-0.000 (0.023)	0.000 (0.022)	-0.000 (0.023)	
Median (Q1, Q3)	0.001 (-0.018, 0.020)	0.002 (-0.018, 0.019)	0.002 (-0.018, 0.019)	
Min - Max	-0.061 - 0.045	-0.070 - 0.046	-0.070 - 0.046	
Missing	0	0	0	
PC2				0.018
Mean (SD)	-0.001 (0.023)	0.001 (0.023)	0.000 (0.023)	
Median (Q1, Q3)	-0.008 (-0.017, 0.009)	-0.004 (-0.016, 0.011)	-0.006 (-0.016, 0.010)	
Min - Max	-0.035 - 0.109	-0.034 - 0.099	-0.035 - 0.109	
Missing	0	0	0	
PC3				0.080
Mean (SD)	0.001 (0.022)	-0.001 (0.023)	-0.000 (0.023)	
Median (Q1, Q3)	0.002 (-0.013, 0.016)	0.001 (-0.016, 0.015)	0.001 (-0.015, 0.016)	
Min - Max	-0.073 - 0.069	-0.073 - 0.065	-0.073 - 0.069	
Missing	0	0	0	
PC4				0.367
Mean (SD)	-0.001 (0.022)	0.000 (0.023)	0.000 (0.023)	
Median (Q1, Q3)	-0.001 (-0.015, 0.015)	0.001 (-0.015, 0.016)	0.000 (-0.015, 0.015)	
Min - Max	-0.069 - 0.054	-0.082 - 0.066	-0.082 - 0.066	
Missing	0	0	0	
PC5				0.088
Mean (SD)	-0.001 (0.022)	0.001 (0.023)	-0.000 (0.023)	
Median (Q1, Q3)	-0.001 (-0.016, 0.013)	0.000 (-0.014, 0.016)	-0.000 (-0.014, 0.014)	
Min - Max	-0.080 - 0.070	-0.080 - 0.098	-0.080 - 0.098	
Missing	0	0	0	
PC6				0.224
Mean (SD)	0.001 (0.021)	-0.001 (0.023)	0.000 (0.023)	
Median (Q1, Q3)	0.002 (-0.010, 0.015)	0.002 (-0.012, 0.014)	0.002 (-0.011, 0.014)	
Min - Max	-0.112 - 0.054	-0.122 - 0.057	-0.122 - 0.057	

	1 (N=702)	2 (N=1241)	Total (N=1943)	p value
Missing	0	0	0	
PC7				0.208
Mean (SD)	0.001 (0.023)	-0.000 (0.022)	0.000 (0.023)	
Median (Q1, Q3)	0.004 (-0.008, 0.015)	0.003 (-0.009, 0.014)	0.004 (-0.009, 0.014)	
Min - Max	-0.121 - 0.053	-0.115 - 0.052	-0.121 - 0.053	
Missing	0	0	0	
PC8				0.787
Mean (SD)	0.000 (0.023)	-0.000 (0.023)	0.000 (0.023)	
Median (Q1, Q3)	0.001 (-0.015, 0.016)	0.001 (-0.015, 0.015)	0.001 (-0.015, 0.016)	
Min - Max	-0.072 - 0.075	-0.071 - 0.076	-0.072 - 0.076	
Missing	0	0	0	
PC9				0.540
Mean (SD)	-0.000 (0.023)	0.000 (0.023)	-0.000 (0.023)	
Median (Q1, Q3)	-0.001 (-0.015, 0.015)	-0.000 (-0.013, 0.013)	-0.001 (-0.014, 0.014)	
Min - Max	-0.068 - 0.106	-0.072 - 0.086	-0.072 - 0.106	
Missing	0	0	0	
PC10				0.491
Mean (SD)	-0.000 (0.023)	0.000 (0.023)	-0.000 (0.023)	
Median (Q1, Q3)	-0.000 (-0.015, 0.016)	0.001 (-0.015, 0.016)	0.000 (-0.015, 0.016)	
Min - Max	-0.075 - 0.072	-0.077 - 0.077	-0.077 - 0.077	
Missing	0	0	0	

```
bpmed_table <- tableby(bpmed_fmla, data = phenotype, control=my_controls)
summary(bpmed_table, labelTranslations = my_labels, title = "Summary Statistics for Phenotype Data By Bl
```

Table 3: Summary Statistics for Phenotype Data By Blood Pressure Medication Status

	0 (N=260)	1 (N=147)	Total (N=407)	p value
Diastolic Blood Pressure				0.061
Mean (SD)	73.303 (11.161)	74.711 (11.126)	73.811 (11.155)	
Median (Q1, Q3)	73.000 (66.000, 80.625)	76.000 (67.500, 83.000)	74.000 (67.000, 81.750)	
Min - Max	49.000 - 115.000	42.000 - 109.000	42.000 - 115.000	
Missing	0	0	0	
Systolic Blood Pressure				< 0.001
Mean (SD)	126.839 (18.159)	145.367 (21.082)	133.531 (21.203)	
Median (Q1, Q3)	125.000 (114.000, 136.125)	143.000 (132.750, 156.000)	131.000 (118.000, 145.000)	
Min - Max	91.000 - 197.000	100.000 - 212.000	91.000 - 212.000	
Missing	0	0	0	
Age				< 0.001
Mean (SD)	50.476 (13.539)	61.490 (12.521)	54.454 (14.191)	
Median (Q1, Q3)	48.990 (42.000, 59.000)	62.248 (52.000, 71.647)	53.000 (44.270, 64.703)	
Min - Max	19.000 - 88.630	35.000 - 91.970	19.000 - 91.970	

	0 (N=260)	1 (N=147)	Total (N=407)	p value
Missing	0	0	0	
PC1				0.966
Mean (SD)	-0.014 (0.019)	-0.014 (0.018)	-0.014 (0.019)	
Median (Q1, Q3)	-0.015 (-0.027, -0.003)	-0.015 (-0.026, -0.004)	-0.015 (-0.026, -0.003)	
Min - Max	-0.061 - 0.045	-0.052 - 0.040	-0.061 - 0.045	
Missing	0	0	0	
PC2				0.917
Mean (SD)	-0.014 (0.006)	-0.014 (0.007)	-0.014 (0.006)	
Median (Q1, Q3)	-0.015 (-0.018, -0.011)	-0.015 (-0.018, -0.011)	-0.015 (-0.018, -0.011)	
Min - Max	-0.032 - 0.005	-0.029 - 0.015	-0.032 - 0.015	
Missing	0	0	0	
PC3				0.659
Mean (SD)	0.003 (0.021)	0.003 (0.022)	0.003 (0.022)	
Median (Q1, Q3)	0.004 (-0.011, 0.018)	0.003 (-0.012, 0.017)	0.003 (-0.012, 0.018)	
Min - Max	-0.056 - 0.055	-0.059 - 0.059	-0.059 - 0.059	
Missing	0	0	0	
PC4				< 0.001
Mean (SD)	0.004 (0.022)	-0.003 (0.022)	0.002 (0.022)	
Median (Q1, Q3)	0.006 (-0.009, 0.021)	-0.004 (-0.017, 0.013)	0.003 (-0.013, 0.018)	
Min - Max	-0.061 - 0.053	-0.055 - 0.054	-0.061 - 0.054	
Missing	0	0	0	
PC5				0.416
Mean (SD)	0.001 (0.021)	-0.001 (0.022)	0.000 (0.021)	
Median (Q1, Q3)	0.001 (-0.012, 0.016)	0.002 (-0.016, 0.014)	0.001 (-0.013, 0.015)	
Min - Max	-0.072 - 0.049	-0.060 - 0.056	-0.072 - 0.056	
Missing	0	0	0	
PC6				0.715
Mean (SD)	0.002 (0.017)	0.002 (0.018)	0.002 (0.017)	
Median (Q1, Q3)	0.001 (-0.009, 0.012)	0.003 (-0.011, 0.014)	0.002 (-0.010, 0.013)	
Min - Max	-0.041 - 0.051	-0.033 - 0.042	-0.041 - 0.051	
Missing	0	0	0	
PC7				0.407
Mean (SD)	0.006 (0.018)	0.003 (0.021)	0.005 (0.019)	
Median (Q1, Q3)	0.007 (-0.004, 0.018)	0.006 (-0.007, 0.016)	0.007 (-0.004, 0.017)	
Min - Max	-0.063 - 0.053	-0.081 - 0.046	-0.081 - 0.053	
Missing	0	0	0	
PC8				0.307
Mean (SD)	-0.000 (0.022)	-0.003 (0.023)	-0.001 (0.023)	
Median (Q1, Q3)	0.000 (-0.016, 0.014)	-0.002 (-0.015, 0.011)	-0.001 (-0.016, 0.013)	
Min - Max	-0.061 - 0.076	-0.071 - 0.055	-0.071 - 0.076	
Missing	0	0	0	
PC9				0.219
Mean (SD)	0.000 (0.019)	-0.003 (0.022)	-0.001 (0.020)	
Median (Q1, Q3)	0.000 (-0.013, 0.013)	-0.003 (-0.016, 0.009)	-0.001 (-0.014, 0.012)	
Min - Max	-0.046 - 0.066	-0.056 - 0.048	-0.056 - 0.066	
Missing	0	0	0	
PC10				0.922
Mean (SD)	0.003 (0.022)	0.003 (0.021)	0.003 (0.022)	
Median (Q1, Q3)	0.003 (-0.012, 0.019)	0.004 (-0.010, 0.017)	0.004 (-0.012, 0.018)	

	0 (N=260)	1 (N=147)	Total (N=407)	p value
Min - Max	-0.060 - 0.077	-0.054 - 0.052	-0.060 - 0.077	
Missing	0	0	0	

```
gen_table <- tableby(~DBP + SBP + AGE, data=phenotype, control=my_controls)
summary(gen_table, labelTranslations = my_labels, title="Summary Statistics for Phenotype Data")
```

Table 4: Summary Statistics for Phenotype Data

	Overall (N=1943)
Diastolic Blood Pressure	
Mean (SD)	73.469 (11.022)
Median (Q1, Q3)	73.000 (66.000, 80.000)
Min - Max	32.000 - 123.000
Missing	92
Systolic Blood Pressure	
Mean (SD)	125.044 (21.103)
Median (Q1, Q3)	122.000 (110.000, 137.000)
Min - Max	66.000 - 213.000
Missing	92
Age	
Mean (SD)	49.485 (14.243)
Median (Q1, Q3)	48.000 (38.000, 60.000)
Min - Max	19.000 - 92.000
Missing	81

```
summary(freqlist(~SEX, data=phenotype), title="Frequency Table By Sex for Phenotype Data", labelTransla
```

Table 5: Frequency Table By Sex for Phenotype Data

Sex	Freq	Cumulative Freq	Percent	Cumulative Percent
1	702	702	36.13	36.13
2	1241	1943	63.87	100.00

```
summary(freqlist(~BPMED, data=phenotype), title="Frequency Table By Blood Pressure Medication Status fo
```

Table 6: Frequency Table By Blood Pressure Medication Status for Phenotype Data

Blood Pressure Medication Status	Freq	Cumulative Freq	Percent	Cumulative Percent
0	260	260	13.38	13.38
1	147	407	7.57	20.95
NA	1536	1943	79.05	100.00

```
summary(freqlist(~HTN, data=phenotype), title="Frequency Table By Hypertension Status for Phenotype Data
```

Table 7: Frequency Table By Hypertension Status for Phenotype Data

Hypertension Status	Freq	Cumulative Freq	Percent	Cumulative Percent
0	1212	1212	62.38	62.38
1	639	1851	32.89	95.27
NA	92	1943	4.73	100.00

Hypertension has significant associations with DBP, SBP, AGE, PC1, PC2.

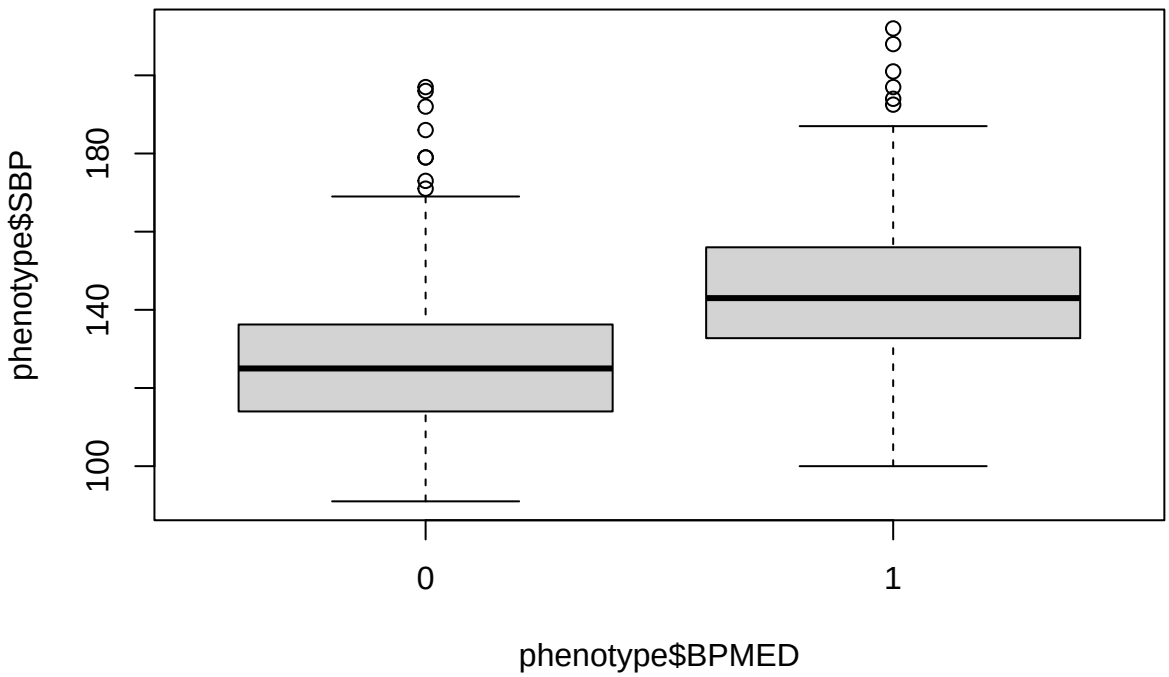
Sex has significant associations with DBP, SBP, AGE, and PC2

Blood Pressure Medication Status has significant assocciations with SBP, AGE, and PC4 (most significant PC yet). It makes sense for it not to have an association with DBP since high blood pressure is due to high SBP. It is interesting to note that hypertension does have a significant association with DBP. It appears that those with hypertension have a higher likelihood of having a high DBP.

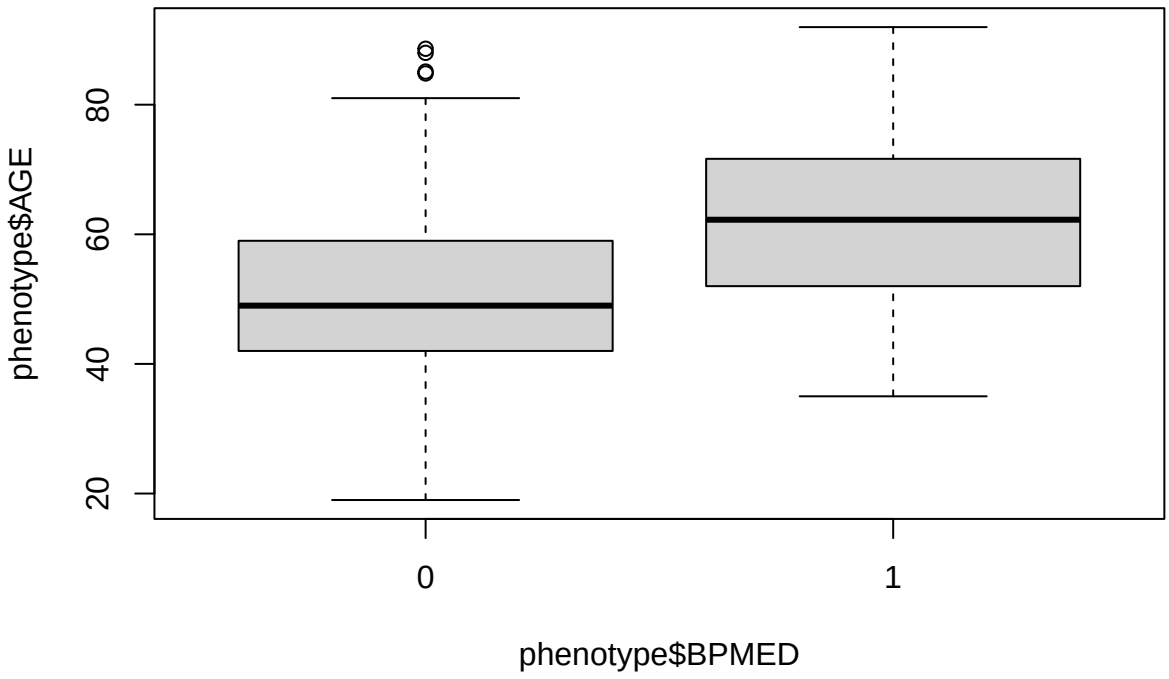
There were only 3 signficant PCs in this analysis - PC1, PC2, and PC4.

There are many more missing BPMED instances (1536) compared to missing SBP and DBP (92). This does not impact hypertension as it was coded to still code depending on SBP if BPMED was null.

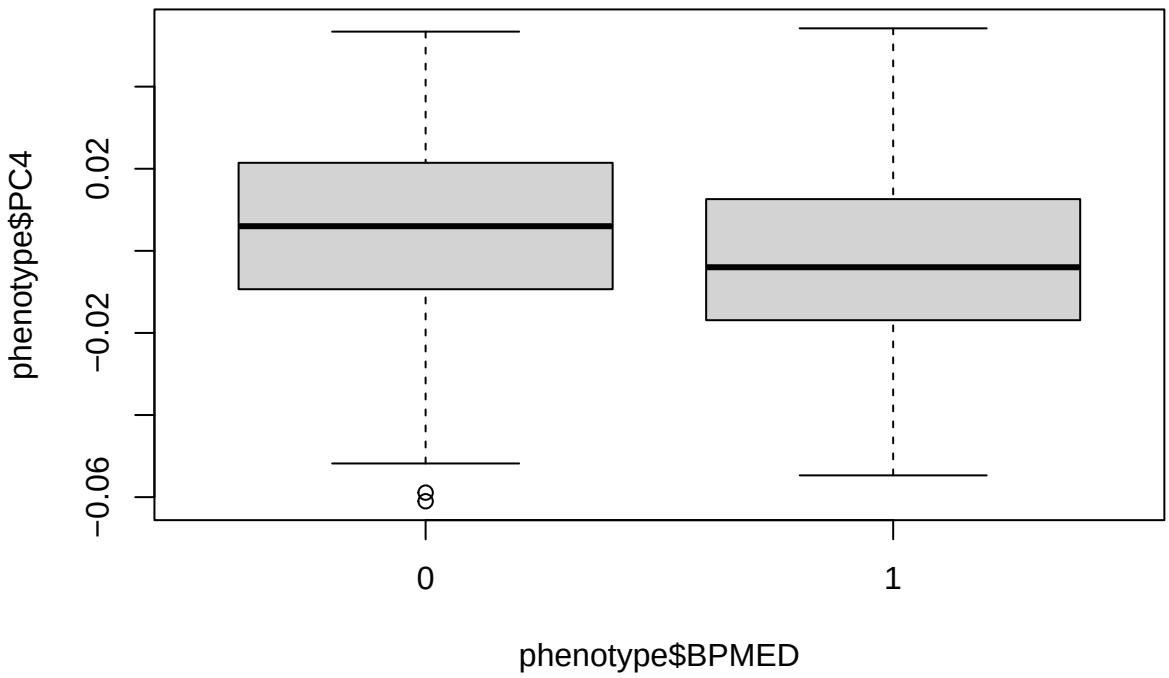
```
boxplot(phenotype$SBP~phenotype$BPMED)
```



```
boxplot(phenotype$AGE~phenotype$BPMED)
```




```
boxplot(phenotype$PC4~phenotype$BPMED)
```




```
bpmed_fmla <- as.formula(paste("BPMED ~ SBP + AGE + SEX + BPMED + HTN + + ", paste(pc_list, collapse="+
bpmed_model <- glm(bpmed_fmla,family=binomial(link='logit'), data=phenotype)
```

```
## Warning in model.matrix.default(mt, mf, contrasts): the response appeared on the
## right-hand side and was dropped
```

```
## Warning in model.matrix.default(mt, mf, contrasts): problem with term 4 in
## model.matrix: no columns are assigned
```

```
#summary(bpmed_model)

bpmed_model_reduc <- glm(BPMED~SBP+AGE+HTN+PC4, family=binomial(link='logit'), data=phenotype)
summary(bpmed_model_reduc)
```

```
##
## Call:
## glm(formula = BPMED ~ SBP + AGE + HTN + PC4, family = binomial(link = "logit"),
##     data = phenotype)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3242  -0.7858  -0.4329   0.8566   2.3539
##
## Coefficients:
```

```
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -6.713073   1.206632  -5.563 2.64e-08 ***
## SBP           0.020317   0.009419   2.157  0.03100 *
## AGE           0.050556   0.009596   5.269 1.37e-07 ***
## HTN           0.997735   0.373270   2.673  0.00752 **
## PC4         -15.366247   5.613926  -2.737  0.00620 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 532.43  on 406  degrees of freedom
## Residual deviance: 412.57  on 402  degrees of freedom
##   (1536 observations deleted due to missingness)
## AIC: 422.57
##
## Number of Fisher Scoring iterations: 4
```

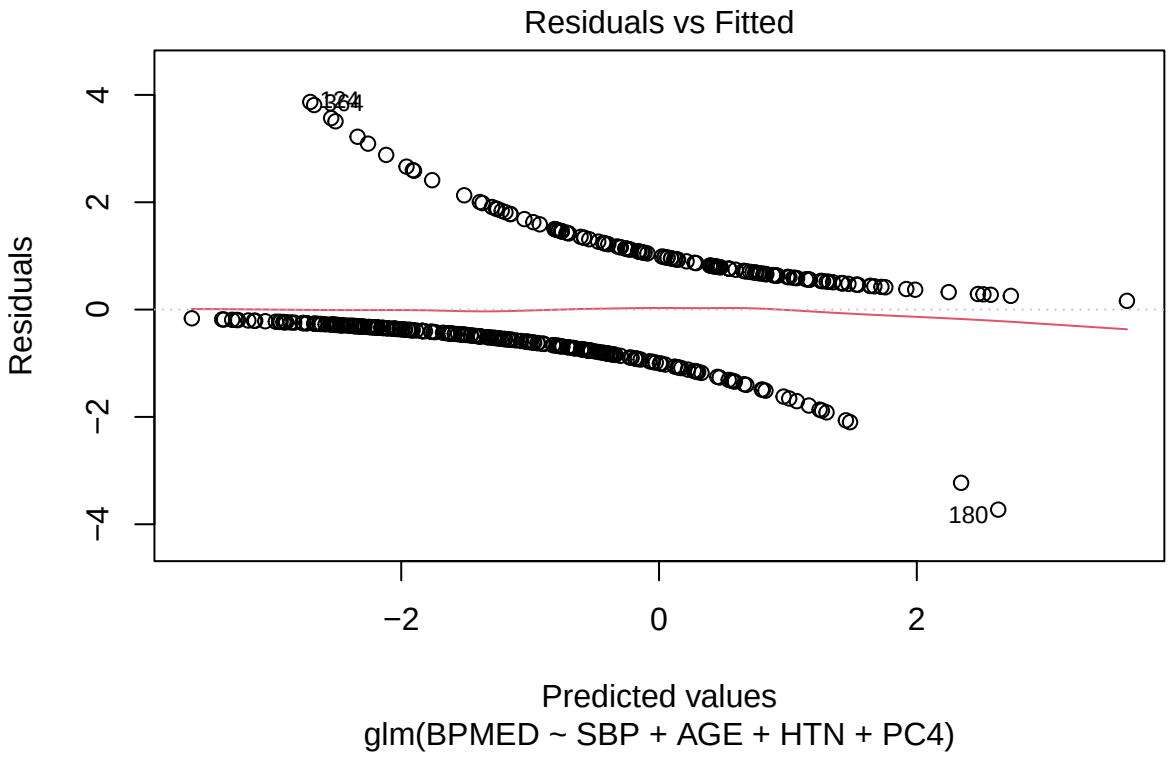
```
anova(bpmed_model_reduc, bpmed_model, test="Chisq")
```

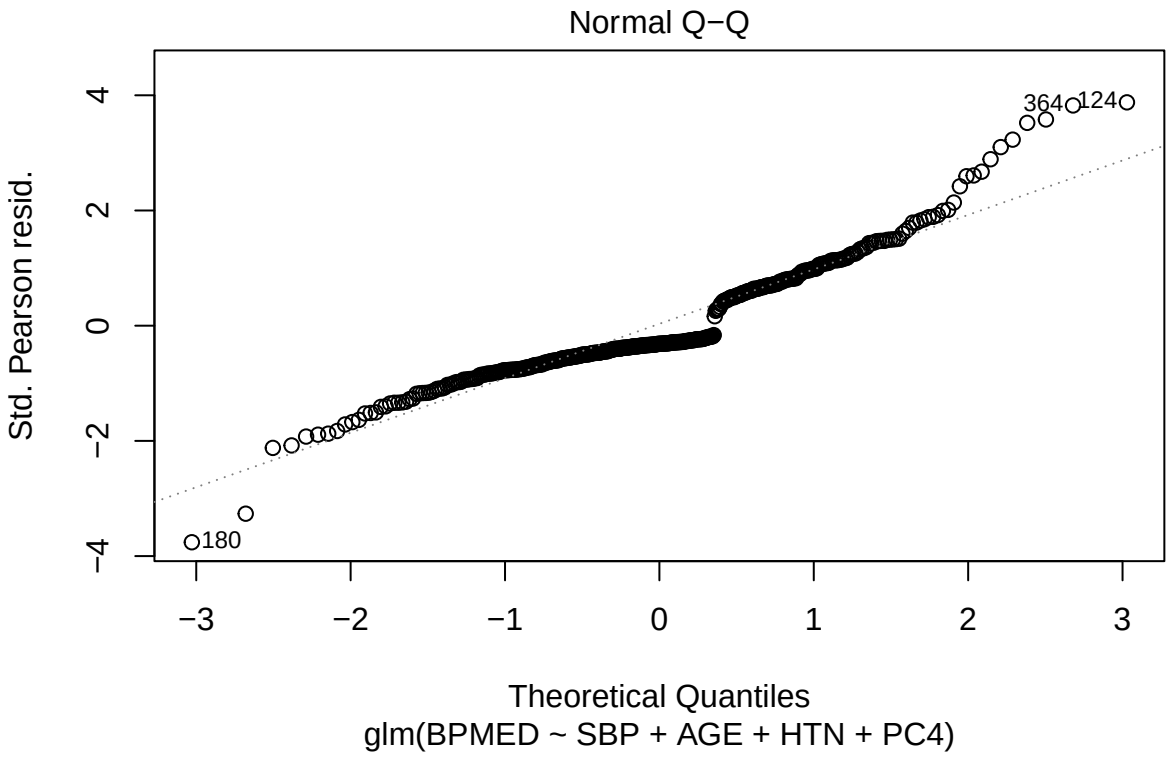
```
## Analysis of Deviance Table
##
## Model 1: BPMED ~ SBP + AGE + HTN + PC4
## Model 2: BPMED ~ SBP + AGE + SEX + BPMED + HTN + +PC1 + PC2 + PC3 + PC4 +
##     PC5 + PC6 + PC7 + PC8 + PC9 + PC10
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       402     412.57
## 2       392     407.59 10   4.9798   0.8925
```

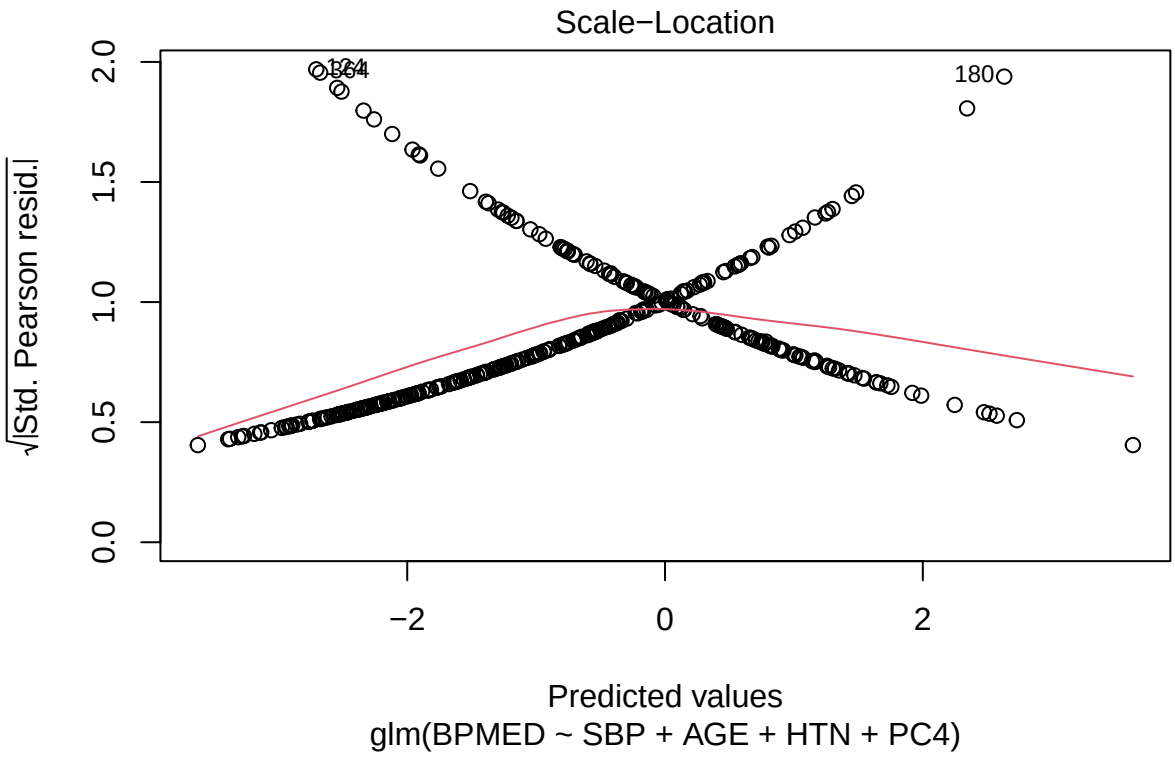
```
exp(coef(bpmed_model_reduc))
```

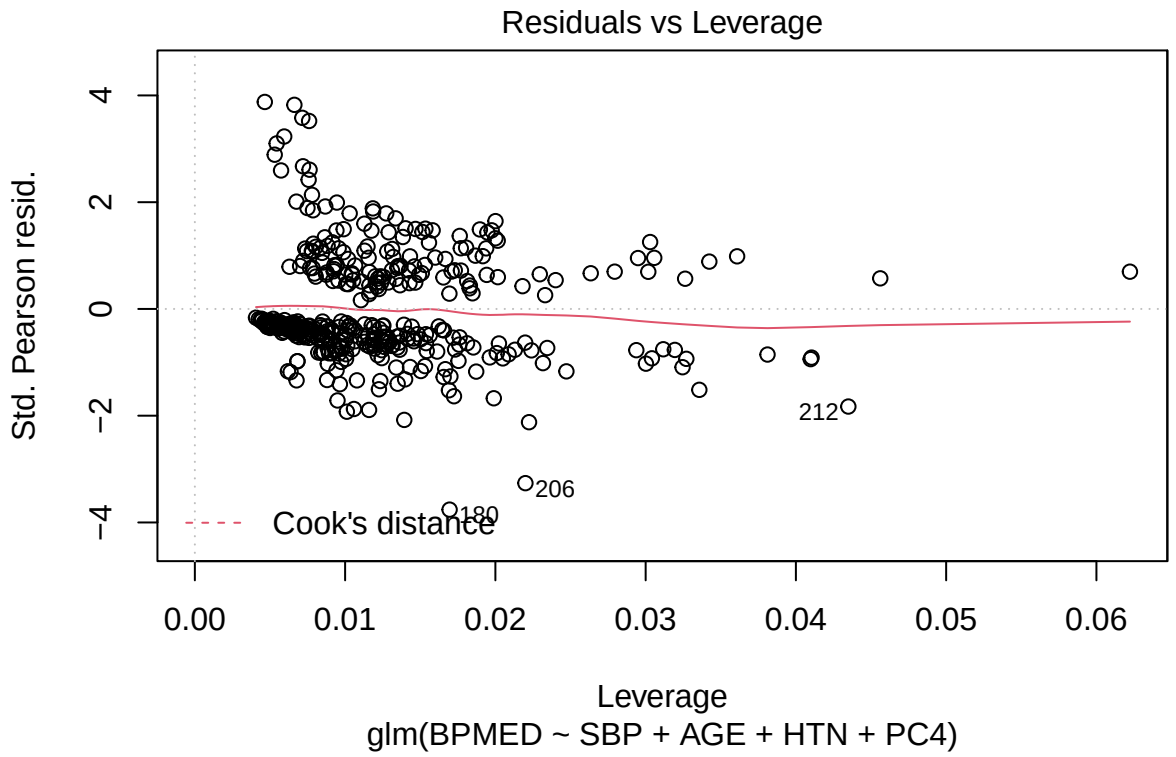
```
##  (Intercept)          SBP          AGE          HTN          PC4
## 1.214925e-03 1.020525e+00 1.051856e+00 2.712132e+00 2.120918e-07
```

```
plot(bpmed_model_reduc) # Little shaky on assumptions of linearity, deviation from normality present, p
```


Residuals vs Fitted
Residuals
Predicted values
glm(BPMED ~ SBP + AGE + HTN + PC4)
180


Normal Q−Q
Std. Pearson resid.
Theoretical Quantiles
glm(BPMED ~ SBP + AGE + HTN + PC4)
180
364
124
−4
−2
0
2
4
−3
−2
−1
0
1
2
3


Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
−2
0
2
180
Predicted values
glm(BPMED ~ SBP + AGE + HTN + PC4)




```
deviance(bpmed_model_reduc)/df.residual(bpmed_model_reduc)
```

```
## [1] 1.02629
```

The reduced model contains the variables SBP, AGE, HTN, and PC4 and appears to be a better fit model than the model with all variables as seen by the p-value of .8925 in the analysis of deviance table. Each unit increase in systolic blood pressure increases the odds of being on blood pressure medication by 1.02 (2%). Each year increase in age increases the odds of being on blood pressure medication by 1.05 (5%). Having hypertension increases the odds of being on blood pressure medication by 2.71 (171%)

It is important to note that logistic regression assumptions may not be met such as homoskedasticity as seen in the scale-location plot and normality as seen in the normal q-q plot. There may also be influential outliers present as seen in the residuals vs leverage plot with 2 points below Cook's distance. It is likely that the influential outliers may come from the SBP variable based on the boxplots.