26_testPRS_trainAoU_valAoU_testMETS735_post-imp_PC-AiR

Heather Wheeler

2024-06-03 13:13:46

```
#load libraries
library(data.table)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.1     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ──────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::between()     masks data.table::between()
## ✖ dplyr::filter()      masks stats::filter()
## ✖ dplyr::first()       masks data.table::first()
## ✖ lubridate::hour()    masks data.table::hour()
## ✖ lubridate::isoweek() masks data.table::isoweek()
## ✖ dplyr::lag()         masks stats::lag()
## ✖ dplyr::last()        masks data.table::last()
## ✖ lubridate::mday()    masks data.table::mday()
## ✖ lubridate::minute()  masks data.table::minute()
## ✖ lubridate::month()   masks data.table::month()
## ✖ lubridate::quarter() masks data.table::quarter()
## ✖ lubridate::second()  masks data.table::second()
## ✖ purrr::transpose()   masks data.table::transpose()
## ✖ lubridate::wday()    masks data.table::wday()
## ✖ lubridate::week()    masks data.table::week()
## ✖ lubridate::yday()    masks data.table::yday()
## ✖ lubridate::year()    masks data.table::year()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#paste function to concatenate filenames/paths
"%&%" = function(a,b) paste(a,b,sep="")
```

```
#read in prs, phenotypes, covariates (file made in 25_compare_AoU_PRS_to_observed_mets735_post-imp.Rmd)
mets = fread("METS735_post-imp_PRSCSx_AoU_scaled_bmi_scores_and_covariates.txt")
#read in mets PCs
pcs = fread("/home/wheelerlab3/2023-09-08_PRSCSx/PRSCSx_testing/mets4prscsx_plink/METS735_PC-AiR_PCs.txt")
mets = left_join(mets,pcs,by=join_by('FID'))
table(mets$country_abbr)
```

```
##
##  GH  JA  SA  US
## 220 176 151 188
```

```
dim(mets)
```

```
## [1] 735  42
```

```
#isa: this is just checking data for outliers

#plot PCs
#ggplot(mets,aes(x=PC1,y=PC2,col=country_abbr)) + geom_point() + scale_color_viridis_d()+ theme_bw(16)
#ggplot(mets,aes(x=PC3,y=PC4,col=country_abbr)) + geom_point() + scale_color_viridis_d()+ theme_bw(16) #rm o
utliers?
#plot age distribution by site
#ggplot(mets,aes(x=age,fill=country_abbr)) + geom_histogram() + scale_fill_viridis_d() + facet_wrap(~sex)

#filter out PC3>0.1 and PC3< -0.1
#filter out PC4>0.1 and PC4< -0.1
#filter out age>120
mets_rm_outliers = filter(mets,PC3>(-0.1) & PC3<0.1,PC4>(-0.1) & PC4<0.1,age<120 )
#ggplot(mets_rm_outliers,aes(x=PC1,y=PC2,col=country_abbr)) + geom_point() + scale_color_viridis_d()+ theme_
bw(16)
#ggplot(mets_rm_outliers,aes(x=PC3,y=PC4,col=country_abbr)) + geom_point() + scale_color_viridis_d()+ theme_
bw(16)
#plot age distribution by site
#ggplot(mets_rm_outliers,aes(x=age,fill=country_abbr)) + geom_histogram() + scale_fill_viridis_d() + facet_w
rap(~sex)

#test with the outliers and missing age removed
mets = mets_rm_outliers
dim(mets)
```

```
## [1] 714  42
```

```
#read in AoU weights post-validation
#aou = fread("mean_aou10fold_pop_weights.txt")
#head(aou)
#calc new score in METS using AoU-scores multiplied by AoU-held-out validation weights
#one new ensemble score per person per AoU validation pop (all,afr,amr,eur)
#make matrix of AoU weights
#aou_mat=as.matrix(aou[,-1])

#make matrix for new ensemble scores
n=714 #mets sample size (outliers and no-age removed)
p=4 #four AoU validation pops
ens_prs = matrix(nrow=n,ncol=p)

for(i in 1:4){
  #need 1x3 matrix of AoU weights

    aou = fread("AoU_scaled_bmi_phi1e-02_popweights_6-3-24.txt")
    aou_mat=as.matrix(aou[,-1])
    aou_weights = t(as.matrix(aou_mat[,i]))

  #need 3x714 matrix of mets PRS's for matrix multiplication
  mets_scores = t(as.matrix(select(mets,AFR,AMR,EUR)))
  #multiply each mets PRS by the AoU weight and take the sum
  # %*% does this via matrix multiplication, makes 1x714 matrix
  new_score = aou_weights %*% mets_scores
  ens_prs[,i] = new_score
}
colnames(ens_prs) = colnames(aou_mat)
rownames(ens_prs) = mets$FID
head(ens_prs)
```

```
##              allAoU       afrAoU       amrAoU       eurAoU
## 4327   0.54105447   0.29057426   0.53853805   0.66320015
## 4542  -0.20584607  -0.19253850  -0.16738876  -0.18507024
## 4075   0.13477449   0.06815577   0.13777541   0.17956970
## 4524   0.39122624   0.26503560   0.37040938   0.47389338
## 4629   0.09823187   0.10466854   0.07624379   0.09206169
## 4220  -0.15034532  -0.10508915  -0.14619376  -0.21373481
```

```
#make df to join with mets
ens_prs_df = as.data.frame(ens_prs) |> rownames_to_column("FID")
#make FID character in mets
mets = mutate(mets,FID=as.character(FID))
all_mets = left_join(ens_prs_df,mets,by="FID")
```

raw correlation

```
ensemble_prs_bmi=select(all_mets,ends_with("AoU"),bmi)
cor(ensemble_prs_bmi,use="pairwise")
```

```
##             allAoU      afrAoU      amrAoU      eurAoU         bmi
## allAoU  1.0000000   0.9327509   0.9945106   0.9879121   0.1622063
## afrAoU  0.9327509   1.0000000   0.8904072   0.8747956   0.1940663
## amrAoU  0.9945106   0.8904072   1.0000000   0.9973052   0.1482495
## eurAoU  0.9879121   0.8747956   0.9973052   1.0000000   0.1406393
## bmi     0.1622063   0.1940663   0.1482495   0.1406393   1.0000000
```

```
cor(ensemble_prs_bmi,use='pairwise',method='spearman')
```

```
##             allAoU      afrAoU      amrAoU      eurAoU         bmi
## allAoU  1.0000000   0.9338572   0.9944218   0.9867646   0.1432626
## afrAoU  0.9338572   1.0000000   0.8935735   0.8781621   0.1658663
## amrAoU  0.9944218   0.8935735   1.0000000   0.9964309   0.1320622
## eurAoU  0.9867646   0.8781621   0.9964309   1.0000000   0.1259066
## bmi     0.1432626   0.1658663   0.1320622   0.1259066   1.0000000
```

```
#does ranking correspond to sample size in AoU?
#next, split by mets country and compare adjust R2
```

Split by METS site and test allAoU, afrAoU, amrAoU, eurAoU

```r
#make matrix to store adjusted R2
mets_res_mat = matrix(nrow=5,ncol=5)
countries = c("GH","SA","JA","US","METS")
for(i in 1:4){
  site = countries[i]
  cat("******",site,'******\n***********************\n')
  data = filter(all_mets,country_abbr==site)
  #baseline model (just covariates)
  cat("******baseline******\n")
#  prs_pcs = select(data,height,sex,age,starts_with("PC")) #all PCs
  prs_pcs = select(data,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
  res = summary(lm(scale(bmi)~.,data=prs_pcs))
  mets_res_mat[i,1] = res$adj.r.squared
  print(res)
  #allAoU ensemble model
  cat("******allAoU******\n")
#  prs_pcs = select(data,allAoU_15k,height,sex,age,starts_with("PC"))
  prs_pcs = select(data,allAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
  res = summary(lm(scale(bmi)~.,data=prs_pcs))
  mets_res_mat[i,2] = res$adj.r.squared
  print(res)
  #afrAoU ensemble model
  cat("******afrAoU******\n")
#  prs_pcs = select(data,afrAoU_5k,height,sex,age,starts_with("PC"))
  prs_pcs = select(data,afrAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
  res = summary(lm(scale(bmi)~.,data=prs_pcs))
  mets_res_mat[i,3] = res$adj.r.squared
  print(res)
  #amrAoU ensemble model
  cat("******amrAoU******\n")
#  prs_pcs = select(data,amrAoU_5k,height,sex,age,starts_with("PC"))
  prs_pcs = select(data,amrAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
  res = summary(lm(scale(bmi)~.,data=prs_pcs))
  mets_res_mat[i,4] = res$adj.r.squared
  print(res)
  #eurAoU ensemble model
  cat("******eurAoU******\n")
#  prs_pcs = select(data,eurAoU_5k,height,sex,age,starts_with("PC"))
  prs_pcs = select(data,eurAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
  res = summary(lm(scale(bmi)~.,data=prs_pcs))
  mets_res_mat[i,5] = res$adj.r.squared
  print(res)
}
```

```
## ****** GH ******
## *************************
## ******baseline******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9983 -0.5295 -0.0682  0.4732  3.9413
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.526540   0.916931  -1.665   0.0975 .
## sex           0.831596   0.134869   6.166 3.64e-09 ***
## age          -0.005602   0.008302  -0.675   0.5006
## PC1         -45.464098  49.363333  -0.921   0.3581
## PC2          21.617487  14.891308   1.452   0.1481
## PC3         -50.941247  25.097471  -2.030   0.0437 *
## PC4         -18.970705  23.067844  -0.822   0.4118
## PC5          40.938763  24.483498   1.672   0.0960 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9028 on 206 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2117, Adjusted R-squared:  0.1849
## F-statistic: 7.902 on 7 and 206 DF,  p-value: 1.748e-08
##
## ******allAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9764 -0.5517 -0.0526  0.4947  3.8732
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.523340   0.918631  -1.658   0.0988 .
## allAoU        0.156326   0.313637   0.498   0.6187
## sex           0.835473   0.135340   6.173 3.53e-09 ***
## age          -0.005641   0.008317  -0.678   0.4984
## PC1         -44.542471  49.488185  -0.900   0.3691
## PC2          21.392114  14.925398   1.433   0.1533
## PC3         -50.673930  25.149098  -2.015   0.0452 *
## PC4         -19.373116  23.124139  -0.838   0.4031
## PC5          40.133727  24.581404   1.633   0.1041
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9045 on 205 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2126, Adjusted R-squared:  0.1819
## F-statistic:  6.92 on 8 and 205 DF,  p-value: 4.652e-08
##
## ******afrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q      Max
## -1.9597 -0.5526 -0.0608  0.4947  3.8909
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.520172   0.918909  -1.654   0.0996 .
## afrAoU         0.250503   0.605060   0.414   0.6793
## sex            0.832767   0.135171   6.161 3.77e-09 ***
## age           -0.005679   0.008321  -0.683   0.4957
## PC1          -45.230401  49.466131  -0.914   0.3616
## PC2           21.569462  14.921798   1.446   0.1498
## PC3          -50.476917  25.173094  -2.005   0.0463 *
## PC4          -19.207441  23.121449  -0.831   0.4071
## PC5           40.569263  24.549116   1.653   0.0999 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9047 on 205 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2123, Adjusted R-squared:  0.1816
## F-statistic: 6.908 on 8 and 205 DF,  p-value: 4.815e-08
##
## ******amrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q      Max
## -1.9821 -0.5513 -0.0528  0.4926  3.8701
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.524142   0.918592  -1.659   0.0986 .
## amrAoU         0.156229   0.305880   0.511   0.6101
## sex            0.836143   0.135405   6.175 3.49e-09 ***
## age           -0.005631   0.008317  -0.677   0.4991
## PC1          -44.379365  49.497714  -0.897   0.3710
## PC2           21.346897  14.927499   1.430   0.1542
## PC3          -50.719206  25.146375  -2.017   0.0450 *
## PC4          -19.389708  23.123896  -0.839   0.4027
## PC5           40.041463  24.590377   1.628   0.1050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9045 on 205 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2127, Adjusted R-squared:  0.182
## F-statistic: 6.922 on 8 and 205 DF,  p-value: 4.626e-08
##
## ******eurAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q      Max
## -1.9849 -0.5498 -0.0531  0.4889  3.8675
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.523708   0.918597  -1.659   0.0987 .
## eurAoU         0.121090   0.237050   0.511   0.6100
## sex            0.836500   0.135453   6.176 3.48e-09 ***
```

```
## age          -0.005637   0.008317  -0.678   0.4987
## PC1         -44.325490  49.502330  -0.895   0.3716
## PC2          21.321832  14.929316   1.428   0.1548
## PC3         -50.665866  25.148392  -2.015   0.0452 *
## PC4         -19.333681  23.120258  -0.836   0.4040
## PC5          40.024902  24.592694   1.628   0.1052
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9045 on 205 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2127, Adjusted R-squared:  0.182
## F-statistic: 6.922 on 8 and 205 DF,  p-value: 4.626e-08
##
## ****** SA ******
## ************************
## ******baseline******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.50435 -0.47145 -0.05511  0.39871  2.71724
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.243236   0.767792  -1.619   0.1077
## sex           1.228188   0.129547   9.481   <2e-16 ***
## age           0.005897   0.009370   0.629   0.5302
## PC1          12.899772   8.481105   1.521   0.1305
## PC2           0.678094   7.362280   0.092   0.9267
## PC3         -20.789675  23.581687  -0.882   0.3795
## PC4         -27.715080  11.612981  -2.387   0.0184 *
## PC5         -44.525934  20.056995  -2.220   0.0280 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7766 on 139 degrees of freedom
## Multiple R-squared:  0.4258, Adjusted R-squared:  0.3969
## F-statistic: 14.73 on 7 and 139 DF,  p-value: 2.864e-14
##
## ******allAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55400 -0.42619 -0.00025  0.38247  2.61243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.959017   0.758216  -1.265   0.20806
## allAoU        0.925451   0.342009   2.706   0.00767 **
## sex           1.242697   0.126812   9.800   < 2e-16 ***
## age           0.002764   0.009237   0.299   0.76519
## PC1          15.639477   8.356142   1.872   0.06338 .
## PC2           2.088236   7.219191   0.289   0.77282
## PC3         -18.236641  23.082331  -0.790   0.43084
## PC4         -28.037789  11.358196  -2.469   0.01479 *
## PC5         -50.027104  19.720939  -2.537   0.01230 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.7595 on 138 degrees of freedom
## Multiple R-squared:  0.4548, Adjusted R-squared:  0.4232
## F-statistic: 14.39 on 8 and 138 DF,  p-value: 3.938e-15
##
## ******afrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.49779 -0.43746 -0.02989  0.37726  2.55171
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.999933   0.753593  -1.327  0.18674
## afrAoU        1.750174   0.612856   2.856  0.00496 **
## sex           1.220854   0.126363   9.662  < 2e-16 ***
## age           0.003549   0.009175   0.387  0.69949
## PC1          15.522281   8.321717   1.865  0.06427 .
## PC2           0.898039   7.180212   0.125  0.90065
## PC3         -16.048304  23.057046  -0.696  0.48758
## PC4         -27.609663  11.325202  -2.438  0.01604 *
## PC5         -50.518977  19.672118  -2.568  0.01129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7573 on 138 degrees of freedom
## Multiple R-squared:  0.4579, Adjusted R-squared:  0.4265
## F-statistic: 14.57 on 8 and 138 DF,  p-value: 2.707e-15
##
## ******amrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.56817 -0.41663  0.00402  0.37422  2.63285
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.958259   0.760126  -1.261   0.2096
## amrAoU        0.886304   0.339693   2.609   0.0101 *
## sex           1.247125   0.127130   9.810   <2e-16 ***
## age           0.002689   0.009263   0.290   0.7721
## PC1          15.598605   8.373430   1.863   0.0646 .
## PC2           2.297864   7.239801   0.317   0.7514
## PC3         -18.984669  23.114324  -0.821   0.4129
## PC4         -28.161757  11.379013  -2.475   0.0145 *
## PC5         -49.790105  19.753985  -2.521   0.0129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7608 on 138 degrees of freedom
## Multiple R-squared:  0.4528, Adjusted R-squared:  0.4211
## F-statistic: 14.28 on 8 and 138 DF,  p-value: 4.968e-15
##
## ******eurAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.56778 -0.43466  0.00612  0.37362  2.65237
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.962332   0.762156  -1.263   0.2088
## eurAoU        0.663476   0.266131   2.493   0.0138 *
## sex           1.247775   0.127426   9.792  <2e-16 ***
## age           0.002745   0.009286   0.296   0.7680
## PC1          15.598548   8.396418   1.858   0.0653 .
## PC2           2.269038   7.256050   0.313   0.7550
## PC3         -19.845135  23.154468  -0.857   0.3929
## PC4         -28.319798  11.403647  -2.483   0.0142 *
## PC5         -49.783837  19.803619  -2.514   0.0131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7624 on 138 degrees of freedom
## Multiple R-squared:  0.4506, Adjusted R-squared:  0.4187
## F-statistic: 14.15 on 8 and 138 DF,  p-value: 6.498e-15
##
## ****** JA ******
## ************************
## ******baseline******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9879 -0.5521 -0.0065  0.4307  3.6803
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.859480   0.745878  -2.493   0.0137 *
## sex           1.158963   0.143686   8.066 1.57e-13 ***
## age          -0.003562   0.008865  -0.402   0.6884
## PC1           4.212483  26.487429   0.159   0.8738
## PC2           1.099920  12.511890   0.088   0.9301
## PC3           1.782624   8.335445   0.214   0.8309
## PC4          -4.790194   8.128785  -0.589   0.5565
## PC5           0.624282   6.119567   0.102   0.9189
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8585 on 161 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.2937, Adjusted R-squared:  0.263
## F-statistic: 9.563 on 7 and 161 DF,  p-value: 6.493e-10
##
## ******allAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0072 -0.4767  0.0016  0.5326  3.6668
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.768022   0.744883  -2.374   0.0188 *
## allAoU        0.434072   0.278128   1.561   0.1206
```

```
## sex               1.161726    0.143060    8.121  1.17e-13 ***
## age              -0.004613    0.008852   -0.521    0.6030
## PC1               1.907901   26.411420    0.072    0.9425
## PC2               0.224771   12.469086    0.018    0.9856
## PC3               0.975943    8.314607    0.117    0.9067
## PC4              -5.455115    8.103987   -0.673    0.5018
## PC5               0.982108    6.096774    0.161    0.8722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8547 on 160 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3043, Adjusted R-squared:  0.2695
## F-statistic: 8.747 on 8 and 160 DF,  p-value: 7e-10
##
## ******afrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9817 -0.5220  0.0033  0.4982  3.6605
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.814211    0.742650   -2.443    0.0157 *
## afrAoU       0.684689    0.422011    1.622    0.1067
## sex          1.149783    0.143075    8.036  1.92e-13 ***
## age         -0.003902    0.008823   -0.442    0.6589
## PC1          3.537216   26.357454    0.134    0.8934
## PC2          0.953308   12.449270    0.077    0.9391
## PC3          0.615578    8.324644    0.074    0.9411
## PC4         -5.148748    8.090907   -0.636    0.5254
## PC5          1.068480    6.094931    0.175    0.8611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8542 on 160 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3051, Adjusted R-squared:  0.2704
## F-statistic: 8.782 on 8 and 160 DF,  p-value: 6.405e-10
##
## ******amrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0111 -0.4757  0.0122  0.5427  3.6661
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.758602    0.745652   -2.358    0.0196 *
## amrAoU       0.437645    0.285007    1.536    0.1266
## sex          1.165570    0.143148    8.142  1.03e-13 ***
## age         -0.004786    0.008864   -0.540    0.5900
## PC1          1.518216   26.434721    0.057    0.9543
## PC2          0.051734   12.478141    0.004    0.9967
## PC3          1.154611    8.310583    0.139    0.8897
## PC4         -5.544152    8.109597   -0.684    0.4952
## PC5          0.970060    6.098080    0.159    0.8738
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8549 on 160 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3039, Adjusted R-squared:  0.2691
## F-statistic: 8.733 on 8 and 160 DF,  p-value: 7.251e-10
##
## ******eurAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0257 -0.4799  0.0132  0.5386  3.6535
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.753389   0.744831  -2.354   0.0198 *
## eurAoU       0.363151   0.221697   1.638   0.1034
## sex          1.169297   0.143080   8.172 8.68e-14 ***
## age         -0.004924   0.008859  -0.556   0.5791
## PC1          1.249953  26.412040   0.047   0.9623
## PC2         -0.035236  12.466271  -0.003   0.9977
## PC3          1.379917   8.295855   0.166   0.8681
## PC4         -5.704319   8.105857  -0.704   0.4826
## PC5          1.078911   6.094150   0.177   0.8597
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.854 on 160 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3053, Adjusted R-squared:  0.2706
## F-statistic: 8.791 on 8 and 160 DF,  p-value: 6.26e-10
##
## ****** US ******
## ************************
## ******baseline******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7578 -0.7416 -0.1056  0.5117  3.6354
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.03117    0.86261  -1.195 0.233593
## sex          0.55874    0.16589   3.368 0.000936 ***
## age         -0.01184    0.01163  -1.018 0.309992
## PC1         34.13421   32.75799   1.042 0.298886
## PC2         16.95108   15.10588   1.122 0.263381
## PC3         -1.27532    7.48935  -0.170 0.864988
## PC4         -0.83713    7.43781  -0.113 0.910520
## PC5         -1.43686    5.40461  -0.266 0.790670
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9717 on 170 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.09322,    Adjusted R-squared:  0.05588
## F-statistic: 2.497 on 7 and 170 DF,  p-value: 0.01821
##
```

```
## ******allAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6823 -0.6601 -0.0981  0.4919  3.5626
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.21304    0.85307  -1.422 0.156877
## allAoU       0.67560    0.27303   2.474 0.014331 *
## sex          0.56709    0.16348   3.469 0.000663 ***
## age         -0.01092    0.01147  -0.952 0.342315
## PC1         37.20856   32.29920   1.152 0.250950
## PC2         16.75109   14.88351   1.125 0.261981
## PC3         -4.82179    7.51690  -0.641 0.522093
## PC4         -0.74934    7.32829  -0.102 0.918678
## PC5         -3.97846    5.42313  -0.734 0.464203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9573 on 169 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1249, Adjusted R-squared:  0.0835
## F-statistic: 3.016 on 8 and 169 DF,  p-value: 0.003434
##
## ******afrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7389 -0.6276 -0.1126  0.5351  3.3165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.10089    0.84283  -1.306  0.19327
## afrAoU       1.11370    0.36709   3.034  0.00280 **
## sex          0.53773    0.16218   3.316  0.00112 **
## age         -0.01127    0.01136  -0.992  0.32262
## PC1         34.65434   31.99550   1.083  0.28031
## PC2         15.14553   14.76605   1.026  0.30650
## PC3         -5.07914    7.42159  -0.684  0.49468
## PC4         -0.79988    7.26459  -0.110  0.91246
## PC5         -4.51759    5.37551  -0.840  0.40187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.949 on 169 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1401, Adjusted R-squared:  0.09935
## F-statistic: 3.441 on 8 and 169 DF,  p-value: 0.001073
##
## ******amrAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6741 -0.6679 -0.1113  0.4761  3.6359
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.22623    0.85780  -1.430 0.154703
## amrAoU       0.63294    0.28923   2.188 0.030018 *
## sex          0.57339    0.16421   3.492 0.000612 ***
## age         -0.01090    0.01151  -0.947 0.344946
## PC1         37.64886   32.43872   1.161 0.247435
## PC2         17.21707   14.94080   1.152 0.250804
## PC3         -4.46812    7.54959  -0.592 0.554751
## PC4         -0.74440    7.35640  -0.101 0.919519
## PC5         -3.58950    5.43513  -0.660 0.509880
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.961 on 169 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1182, Adjusted R-squared:  0.07646
## F-statistic: 2.832 on 8 and 169 DF,  p-value: 0.005647
##
## ******eurAoU******
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6741 -0.6673 -0.1149  0.4769  3.6715
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.22274    0.85999  -1.422 0.156926
## eurAoU       0.46173    0.22743   2.030 0.043902 *
## sex          0.57295    0.16454   3.482 0.000633 ***
## age         -0.01092    0.01153  -0.946 0.345295
## PC1         37.81286   32.51182   1.163 0.246449
## PC2         17.47417   14.97128   1.167 0.244782
## PC3         -4.29433    7.56903  -0.567 0.571225
## PC4         -0.74821    7.37057  -0.102 0.919263
## PC5         -3.27666    5.43178  -0.603 0.547159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9629 on 169 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1148, Adjusted R-squared:  0.07291
## F-statistic:  2.74 on 8 and 169 DF,  p-value: 0.007227
```

```
#add all of METS
data = all_mets
i=5
cat("******baseline******\n")
```

```
## ******baseline******
```

```
prs_pcs = select(data,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
res = summary(lm(scale(bmi)~.,data=prs_pcs))
mets_res_mat[i,1] = res$adj.r.squared
print(res)
```

```
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7988 -0.5515 -0.1122  0.4226  3.8531
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.4467185  0.2374470  -6.093 1.83e-09 ***
## sex          0.9052514  0.0690480  13.110  < 2e-16 ***
## age         -0.0004492  0.0045204  -0.099    0.921
## PC1          4.3133915  1.0224547   4.219 2.78e-05 ***
## PC2         -6.3541290  1.2271470  -5.178 2.94e-07 ***
## PC3         -0.5278048  4.6902408  -0.113    0.910
## PC4         -2.1507250  4.2151447  -0.510    0.610
## PC5         -3.8255479  3.2261952  -1.186    0.236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.865 on 700 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.2593, Adjusted R-squared:  0.2518
## F-statistic:    35 on 7 and 700 DF,  p-value: < 2.2e-16
```

```
#allAoU ensemble model
cat("******allAoU******\n")
```

```
## ******allAoU******
```

```
prs_pcs = select(data,allAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
res = summary(lm(scale(bmi)~.,data=prs_pcs))
mets_res_mat[i,2] = res$adj.r.squared
print(res)
```

```
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8934 -0.5553 -0.0968  0.4234  3.8151
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.419683   0.234010  -6.067 2.14e-09 ***
## allAoU       0.660560   0.140354   4.706 3.04e-06 ***
## sex          0.910976   0.068039  13.389  < 2e-16 ***
## age         -0.001361   0.004458  -0.305    0.760
## PC1          4.536116   1.008461   4.498 8.03e-06 ***
## PC2         -5.706588   1.216822  -4.690 3.29e-06 ***
## PC3         -2.411084   4.638244  -0.520    0.603
## PC4         -2.186560   4.152880  -0.527    0.599
## PC5         -4.533128   3.182087  -1.425    0.155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8522 on 699 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.282,  Adjusted R-squared:  0.2738
## F-statistic: 34.32 on 8 and 699 DF,  p-value: < 2.2e-16
```

```
#afrAoU ensemble model
cat("******afrAoU******\n")
```

```
## ******afrAoU******
```

```
#prs_pcs = select(data,afrAoU,height,sex,age,starts_with("PC"))
prs_pcs = select(data,afrAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
res = summary(lm(scale(bmi)~.,data=prs_pcs))
mets_res_mat[i,3] = res$adj.r.squared
print(res)
```

```
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8492 -0.5552 -0.0827  0.4075  3.5739
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.402013   0.233090  -6.015 2.90e-09 ***
## afrAoU       1.154563   0.216814   5.325 1.36e-07 ***
## sex          0.893611   0.067772  13.185  < 2e-16 ***
## age         -0.001097   0.004436  -0.247    0.805
## PC1          4.609443   1.004582   4.588 5.29e-06 ***
## PC2         -5.809754   1.208181  -4.809 1.86e-06 ***
## PC3         -2.623937   4.617998  -0.568    0.570
## PC4         -1.935701   4.135313  -0.468    0.640
## PC5         -4.639112   3.168628  -1.464    0.144
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8485 on 699 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.2881, Adjusted R-squared:   0.28
## F-statistic: 35.37 on 8 and 699 DF,  p-value: < 2.2e-16
```

```
#amrAoU ensemble model
cat("******amrAoU******\n")
```

```
## ******amrAoU******
```

```
prs_pcs = select(data,amrAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
res = summary(lm(scale(bmi)~.,data=prs_pcs))
mets_res_mat[i,4] = res$adj.r.squared
print(res)
```

```
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8924 -0.5519 -0.0904  0.4275  3.8833
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.425812   0.234454  -6.081 1.96e-09 ***
## amrAoU       0.628698   0.143166   4.391 1.30e-05 ***
## sex          0.915004   0.068200  13.417  < 2e-16 ***
## age         -0.001374   0.004467  -0.307    0.759
## PC1          4.506661   1.010316   4.461 9.52e-06 ***
## PC2         -5.710983   1.220248  -4.680 3.44e-06 ***
## PC3         -2.233858   4.646430  -0.481    0.631
## PC4         -2.238600   4.161198  -0.538    0.591
## PC5         -4.433640   3.187877  -1.391    0.165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8539 on 699 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.2791, Adjusted R-squared:  0.2709
## F-statistic: 33.83 on 8 and 699 DF,  p-value: < 2.2e-16
```

```
#eurAoU ensemble model
cat("******eurAoU******\n")
```

```
## ******eurAoU******
```

```
prs_pcs = select(data,eurAoU,bmi,sex,age,PC1,PC2,PC3,PC4,PC5) #5 PCs
res = summary(lm(scale(bmi)~.,data=prs_pcs))
mets_res_mat[i,5] = res$adj.r.squared
print(res)
```

```
##
## Call:
## lm(formula = scale(bmi) ~ ., data = prs_pcs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8929 -0.5461 -0.0863  0.4263  3.9174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.428994   0.234684  -6.089 1.87e-09 ***
## eurAoU       0.472044   0.111850   4.220 2.76e-05 ***
## sex          0.916417   0.068285  13.421  < 2e-16 ***
## age         -0.001361   0.004472  -0.304    0.761
## PC1          4.498332   1.011344   4.448 1.01e-05 ***
## PC2         -5.698516   1.222581  -4.661 3.77e-06 ***
## PC3         -2.085313   4.649584  -0.448    0.654
## PC4         -2.226844   4.165462  -0.535    0.593
## PC5         -4.274405   3.189912  -1.340    0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8548 on 699 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.2777, Adjusted R-squared:  0.2694
## F-statistic: 33.59 on 8 and 699 DF,  p-value: < 2.2e-16
```

```
rownames(mets_res_mat) = countries
colnames(mets_res_mat) = c("baseline","allAoU","afrAoU","amrAoU","eurAoU")
mets_res_mat
```

```
##        baseline     allAoU     afrAoU    amrAoU     eurAoU
## GH   0.18489007 0.18190534 0.18159822 0.1819549 0.18195519
## SA   0.39692830 0.42316403 0.42645324 0.4211148 0.41873716
## JA   0.26297826 0.26949275 0.27037567 0.2691426 0.27060388
## US   0.05588161 0.08350011 0.09934733 0.0764644 0.07290634
## METS 0.25184835 0.27379047 0.27998757 0.2708930 0.26939468
```

```
table(all_mets$country_abbr)
```

```
##
##  GH  JA  SA  US
## 215 172 147 180
```

```
mets_res_mat/mets_res_mat[,1]
```

```
##      baseline    allAoU    afrAoU    amrAoU    eurAoU
## GH          1 0.9838567 0.9821956 0.9841249 0.9841263
## SA          1 1.0660969 1.0743836 1.0609343 1.0549441
## JA          1 1.0247720 1.0281294 1.0234406 1.0289972
## US          1 1.4942325 1.7778181 1.3683286 1.3046572
## METS        1 1.0871243 1.1117308 1.0756194 1.0696702
```

```
table(all_mets$country_abbr)
```

```
##
##  GH  JA  SA  US
## 215 172 147 180
```

```
mets_res_df = rownames_to_column(as.data.frame(mets_res_mat),"site")
mets_res_df_long = mets_res_df |> pivot_longer(!site, names_to = "PRS_val_pop", values_to = "Adj_R2")
ggplot(mets_res_df_long,aes(x=factor(site,levels=c("GH","SA","JA","US","METS")),y=Adj_R2,fill=factor(PRS_val
_pop,levels=c("baseline","allAoU","afrAoU","amrAoU","eurAoU")))) +
  geom_bar(position="dodge", stat="identity") +
  scale_fill_viridis_d(name="PRS validation\npopulation") +
  labs(x="METS population",y=bquote("Adjusted"~R^2))
```
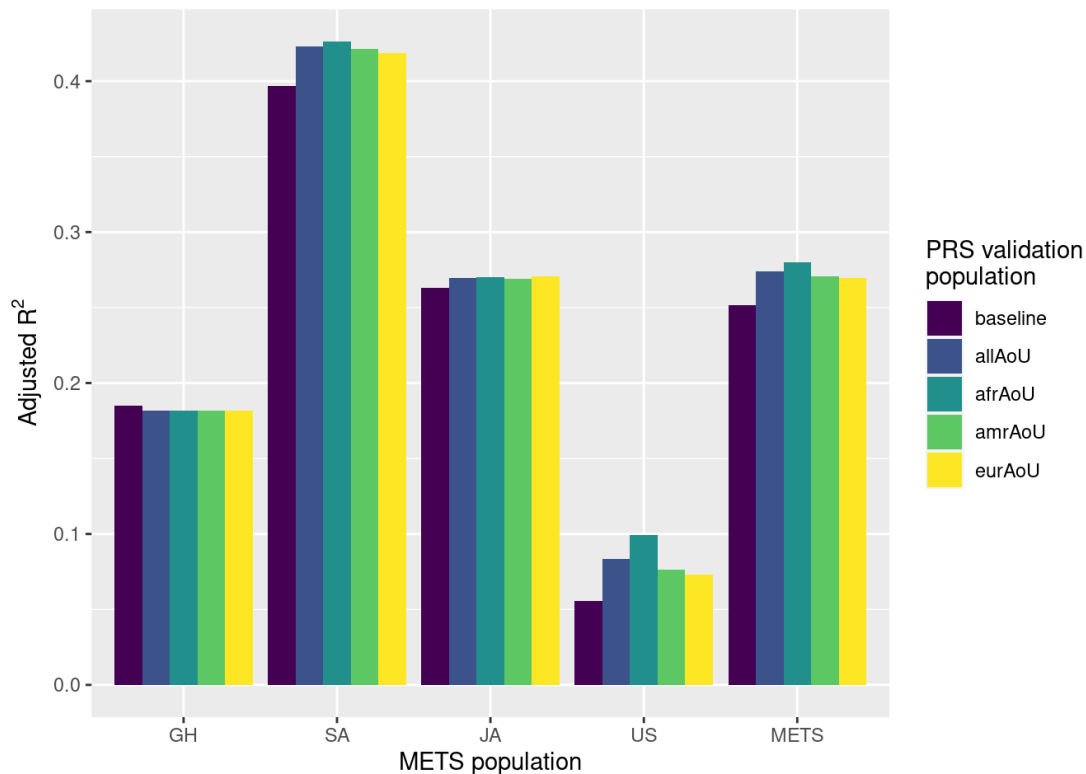


```
#normalize by baseline
norm_mets_res = rownames_to_column(as.data.frame(mets_res_mat/mets_res_mat[,1]),"site")
norm_mets_long = norm_mets_res |> pivot_longer(!site, names_to = "PRS_val_pop", values_to = "Adj_R2")
ggplot(norm_mets_long,aes(x=factor(site,levels=c("GH","SA","JA","US","METS")),y=Adj_R2,fill=factor(PRS_val_p
op,levels=c("baseline","allAoU","afrAoU","amrAoU","eurAoU")))) +
  geom_bar(position="dodge", stat="identity") +
  scale_fill_viridis_d(name="PRS validation\npopulation") +
  labs(x="METS population",y=bquote("Normalized Adjusted"~R^2)) +
  coord_cartesian(ylim=c(0.95,1.85))
```
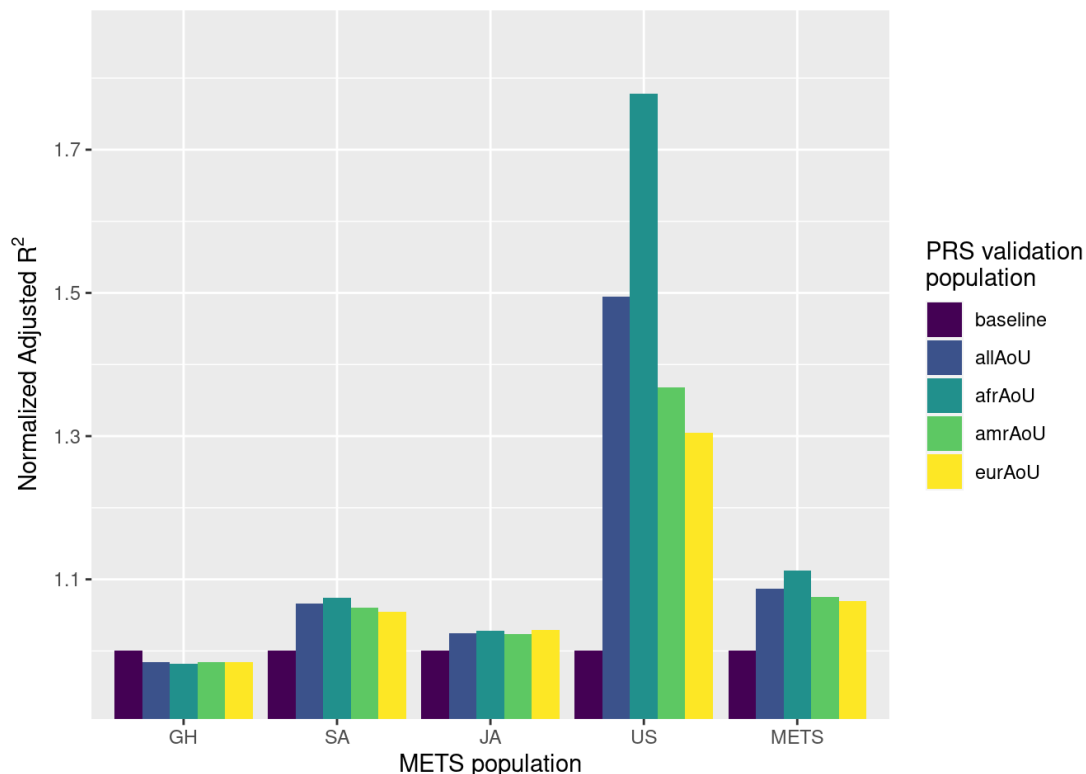
looking at normalized by baseline: makes sense that US is dramatically better because it's closest to aou pops. looking at raw adj r2: i do NOT know

scale across all METS first

```
all_mets=mutate(all_mets,scale_bmi = scale(bmi))
#make matrix to store adjusted R2
mets_res_mat = matrix(nrow=4,ncol=5)
countries = c("GH","SA","JA","US")
for(i in 1:4){
  site = countries[i]
  data = filter(all_mets,country_abbr==site)
  #baseline model (just covariates)
  prs_pcs = select(data,scale_bmi,sex,age,PC1,PC2,PC3,PC4,PC5)
  res = summary(lm(scale_bmi~.,data=prs_pcs))
  mets_res_mat[i,1] = res$adj.r.squared
  #allAoU ensemble model
  prs_pcs = select(data,allAoU,scale_bmi,sex,age,PC1,PC2,PC3,PC4,PC5)
  res = summary(lm(scale_bmi~.,data=prs_pcs))
  mets_res_mat[i,2] = res$adj.r.squared
  #afrAoU ensemble model
  prs_pcs = select(data,afrAoU,scale_bmi,sex,age,PC1,PC2,PC3,PC4,PC5)
  res = summary(lm(scale_bmi~.,data=prs_pcs))
  mets_res_mat[i,3] = res$adj.r.squared
  #amrAoU ensemble model
  prs_pcs = select(data,amrAoU,scale_bmi,sex,age,PC1,PC2,PC3,PC4,PC5)
  res = summary(lm(scale_bmi~.,data=prs_pcs))
  mets_res_mat[i,4] = res$adj.r.squared
  #eurAoU ensemble model
  prs_pcs = select(data,eurAoU,scale_bmi,sex,age,PC1,PC2,PC3,PC4,PC5)
  res = summary(lm(scale_bmi~.,data=prs_pcs))
  mets_res_mat[i,5] = res$adj.r.squared
}
rownames(mets_res_mat) = countries
colnames(mets_res_mat) = c("baseline","allAoU","afrAoU","amrAoU","eurAoU")
mets_res_mat
```

```
##      baseline    allAoU    afrAoU    amrAoU    eurAoU
## GH 0.18489007 0.18190534 0.18159822 0.1819549 0.18195519
## SA 0.39692830 0.42316403 0.42645324 0.4211148 0.41873716
## JA 0.26297826 0.26949275 0.27037567 0.2691426 0.27060388
## US 0.05588161 0.08350011 0.09934733 0.0764644 0.07290634
```

```
mets_res_mat/mets_res_mat[,1]
```

```
##      baseline    allAoU    afrAoU    amrAoU    eurAoU
## GH         1 0.9838567 0.9821956 0.9841249 0.9841263
## SA         1 1.0660969 1.0743836 1.0609343 1.0549441
## JA         1 1.0247720 1.0281294 1.0234406 1.0289972
## US         1 1.4942325 1.7778181 1.3683286 1.3046572
```

```
table(all_mets$country_abbr)
```

```
##
##  GH  JA  SA  US
## 215 172 147 180
```

raw correlation b/t PRS and height

```
#used these numbers (afrAoU) for ASHG abstract
all_mets |> group_by(country_abbr) |> summarize(cor=cor(scale(bmi), allAoU, use="pairwise",method='spearma
n'))
```

```
## # A tibble: 4 × 2
##   country_abbr cor[,1]
##   <chr>          <dbl>
## 1 GH            0.0209
## 2 JA            0.119
## 3 SA            0.0895
## 4 US            0.178
```

```
all_mets |> group_by(country_abbr) |> summarize(cor=cor(scale(bmi), afrAoU, use="pairwise",method='spearma
n'))
```

```
## # A tibble: 4 × 2
##   country_abbr cor[,1]
##   <chr>          <dbl>
## 1 GH            0.0206
## 2 JA            0.141
## 3 SA            0.148
## 4 US            0.220
```

```
all_mets |> group_by(country_abbr) |> summarize(cor=cor(scale(bmi), eurAoU, use="pairwise",method='spearma
n'))
```

```
## # A tibble: 4 × 2
##   country_abbr cor[,1]
##   <chr>          <dbl>
## 1 GH            0.0174
## 2 JA            0.109
## 3 SA            0.0643
## 4 US            0.147
```

plot BMI and PRS distributions

```
#height by site
ggplot(all_mets,aes(x=bmi,col=country_abbr)) + geom_density() + scale_color_viridis_d()
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_density()`).
```



```
#scaled height by site
ggplot(all_mets,aes(x=scale(bmi),col=country_abbr)) + geom_density() + scale_color_viridis_d()
```
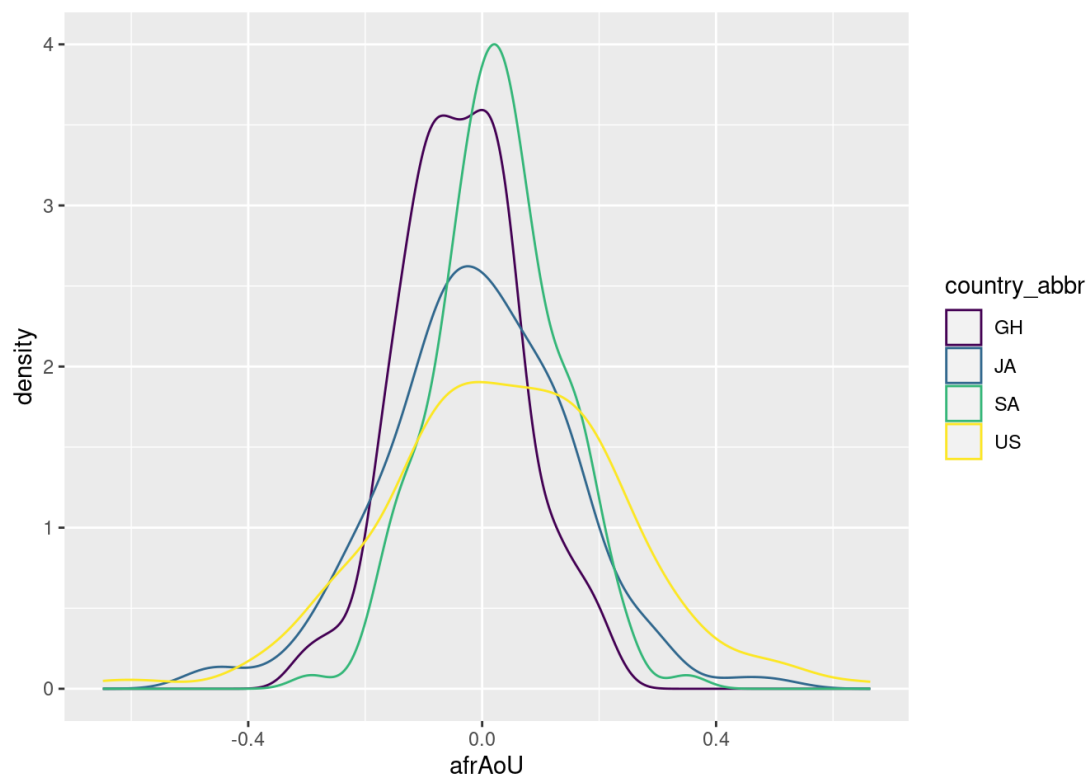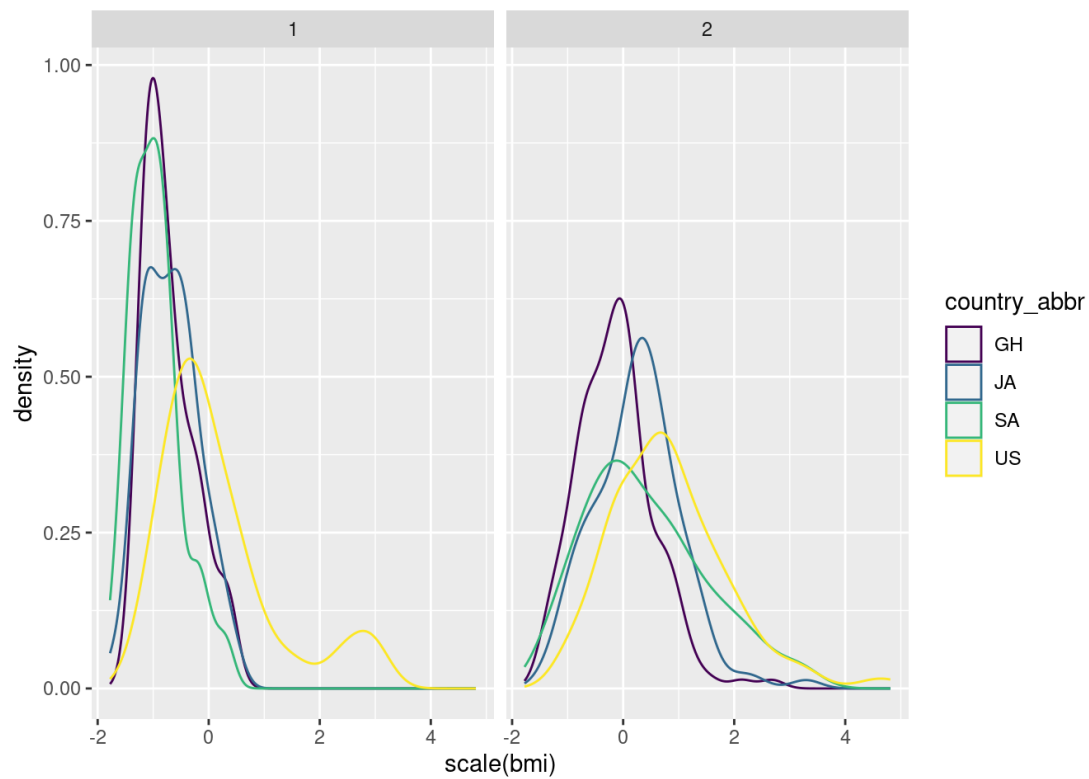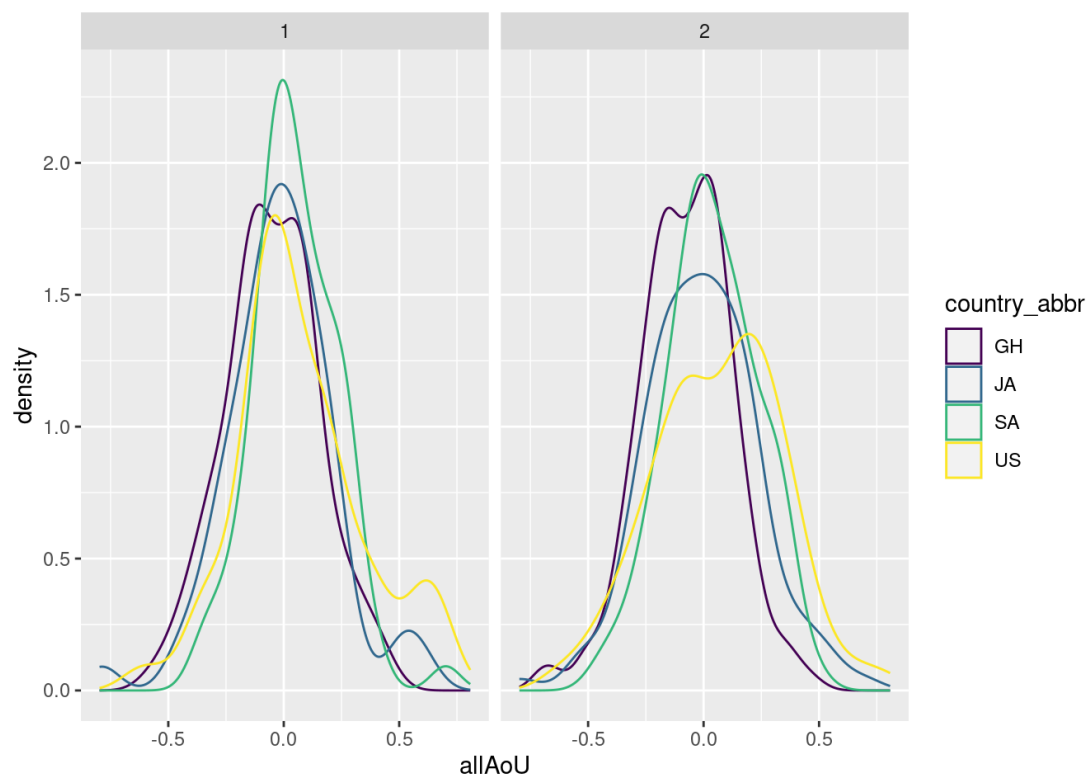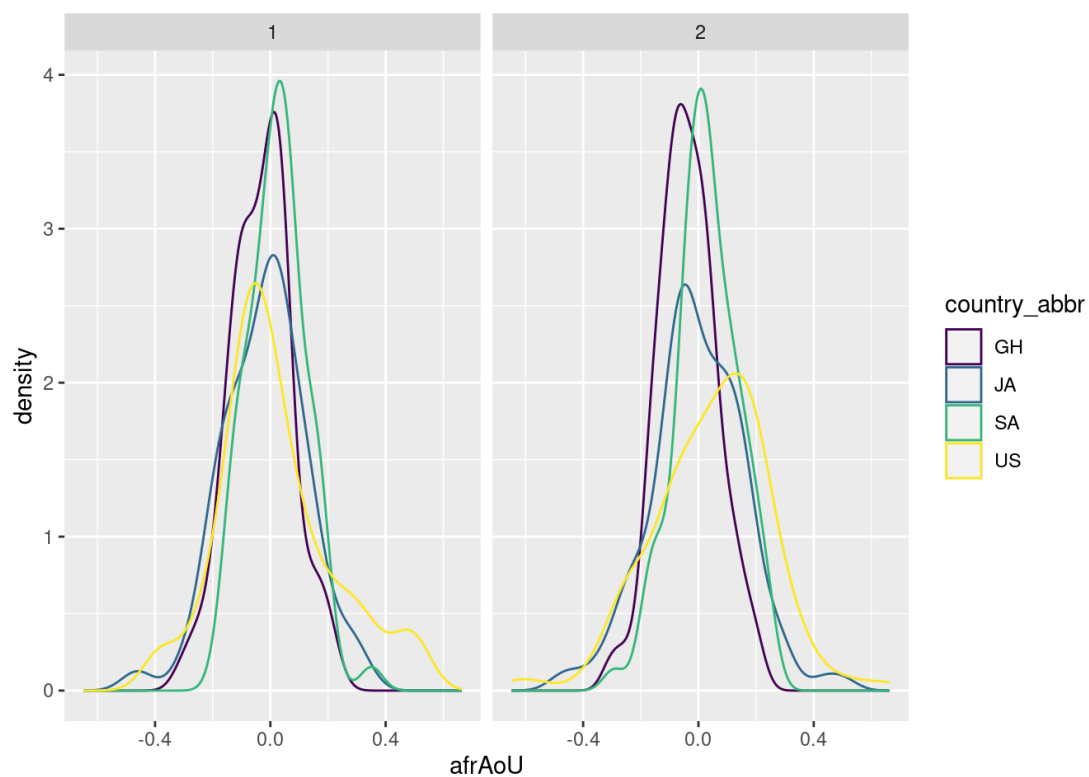
```
## Warning: Removed 6 rows containing non-finite values (`stat_density()`).
```
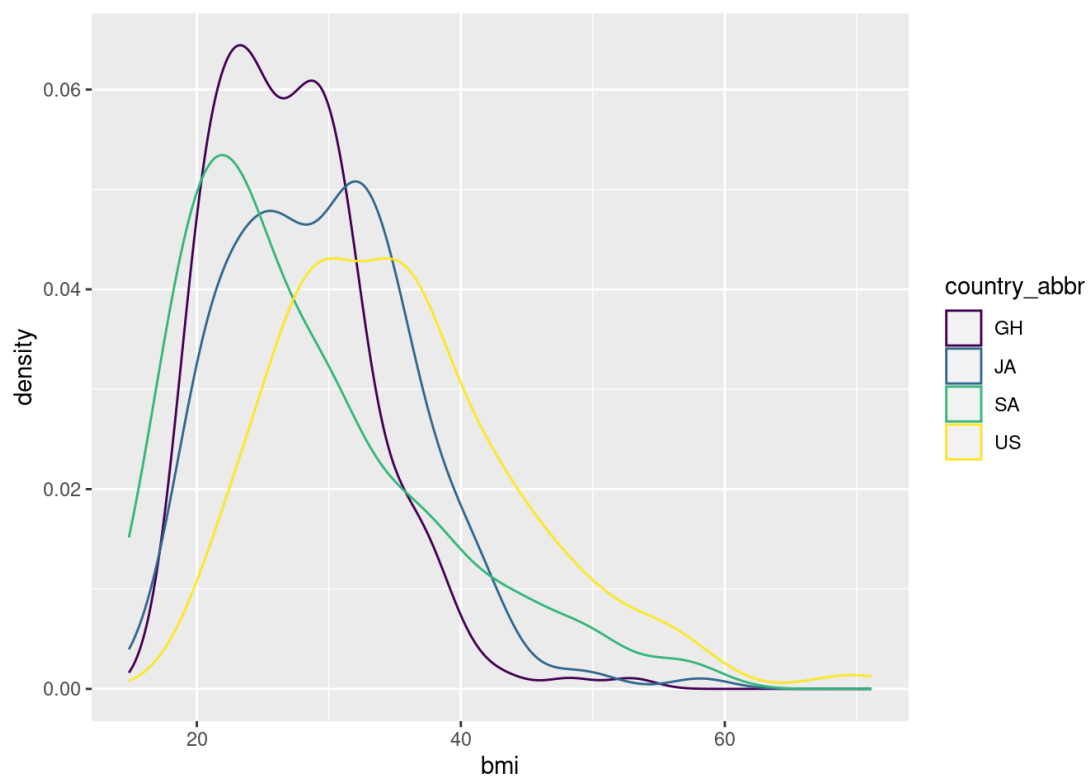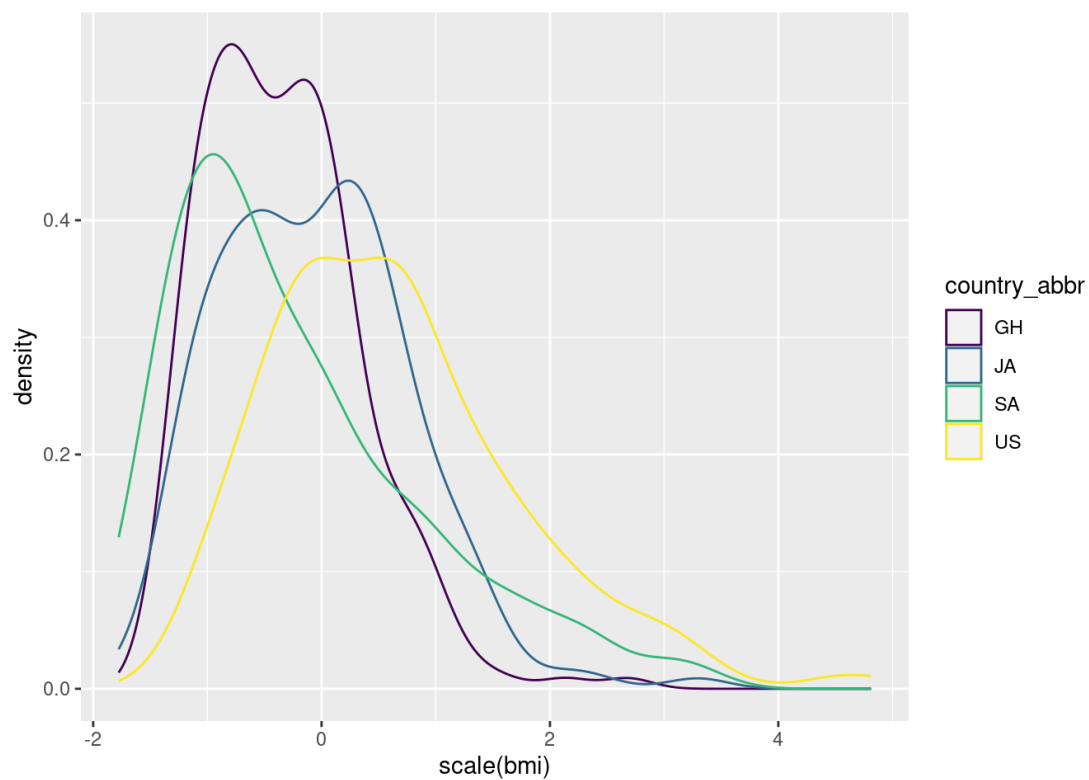
```
#allAoU PRS
ggplot(all_mets,aes(x=allAoU,col=country_abbr)) + geom_density() + scale_color_viridis_d()
```



```
#afrAoU PRS
ggplot(all_mets,aes(x=afrAoU,col=country_abbr)) + geom_density() + scale_color_viridis_d()
```
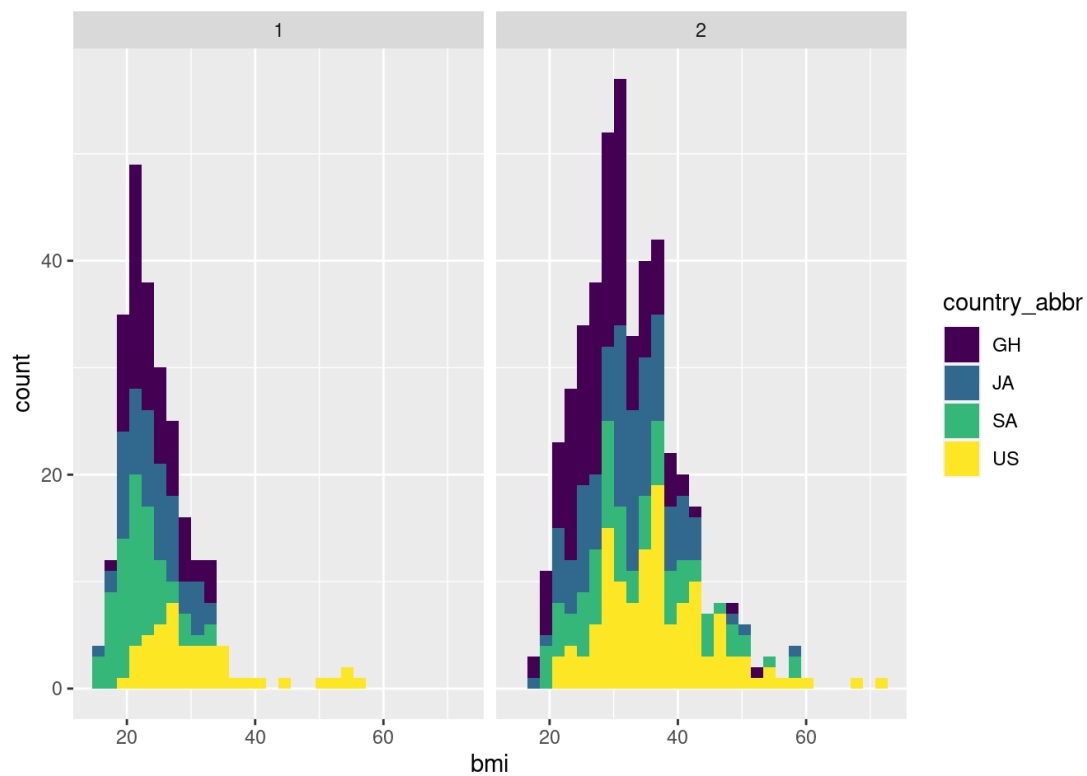
```
#facet by sex
#height by site
ggplot(all_mets,aes(x=bmi,col=country_abbr)) + geom_density() + scale_color_viridis_d() + facet_wrap(~sex)
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_density()`).
```

```
#scaled height by site
ggplot(all_mets,aes(x=scale(bmi),col=country_abbr)) + geom_density() + scale_color_viridis_d() + facet_wrap
(~sex)
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_density()`).
```
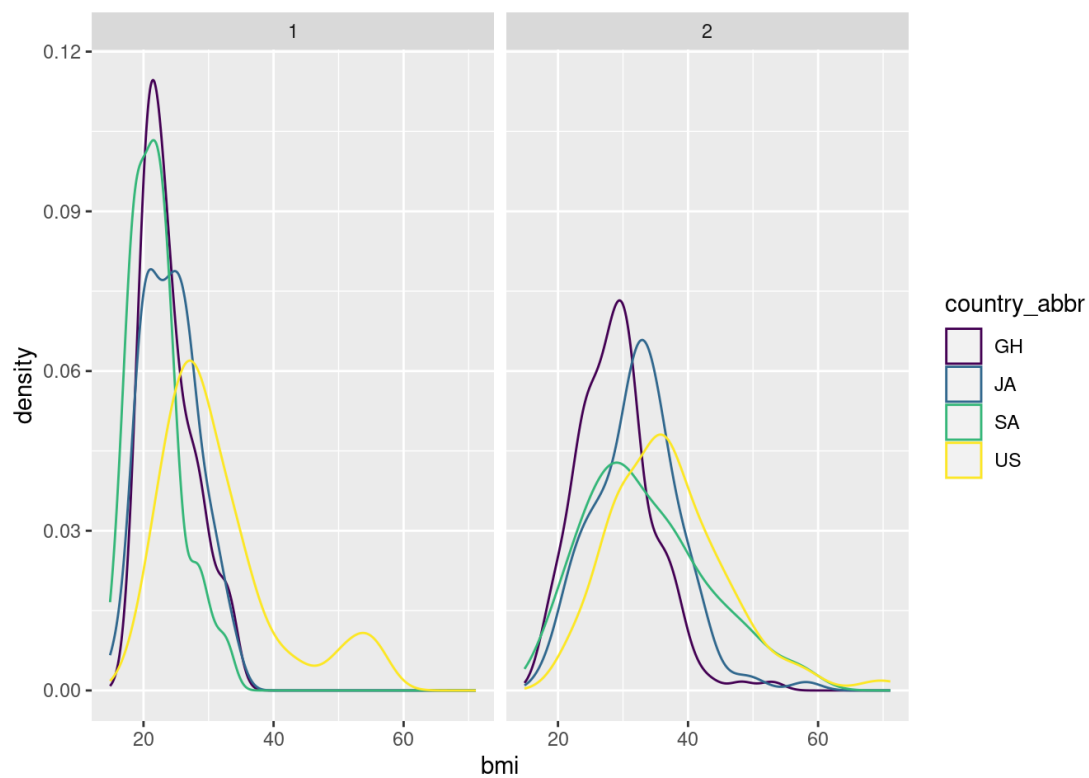


```
#allAoU PRS
ggplot(all_mets,aes(x=allAoU,col=country_abbr)) + geom_density() + scale_color_viridis_d() + facet_wrap(~se
x)
```

```
#afrAoU PRS
ggplot(all_mets,aes(x=afrAoU,col=country_abbr)) + geom_density() + scale_color_viridis_d() + facet_wrap(~se
x)
```



```
#histograms
#height by site
ggplot(all_mets,aes(x=bmi,fill=country_abbr)) + geom_histogram() + scale_fill_viridis_d() + facet_wrap(~sex)
```
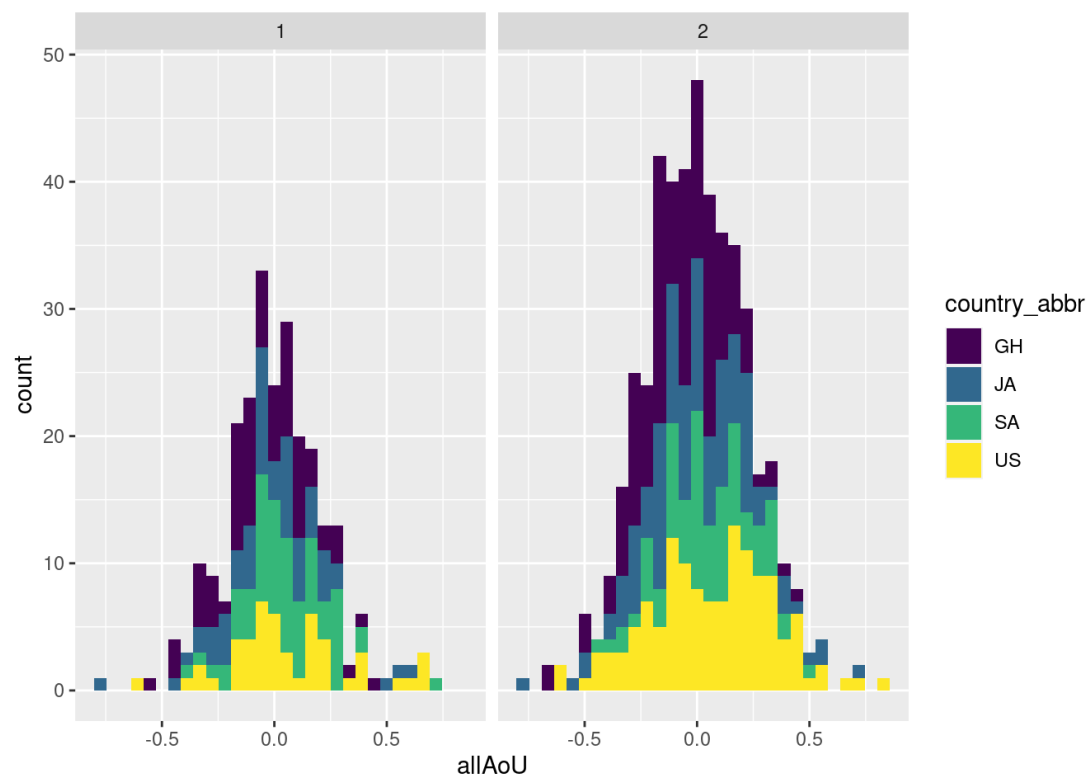
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_bin()`).
```
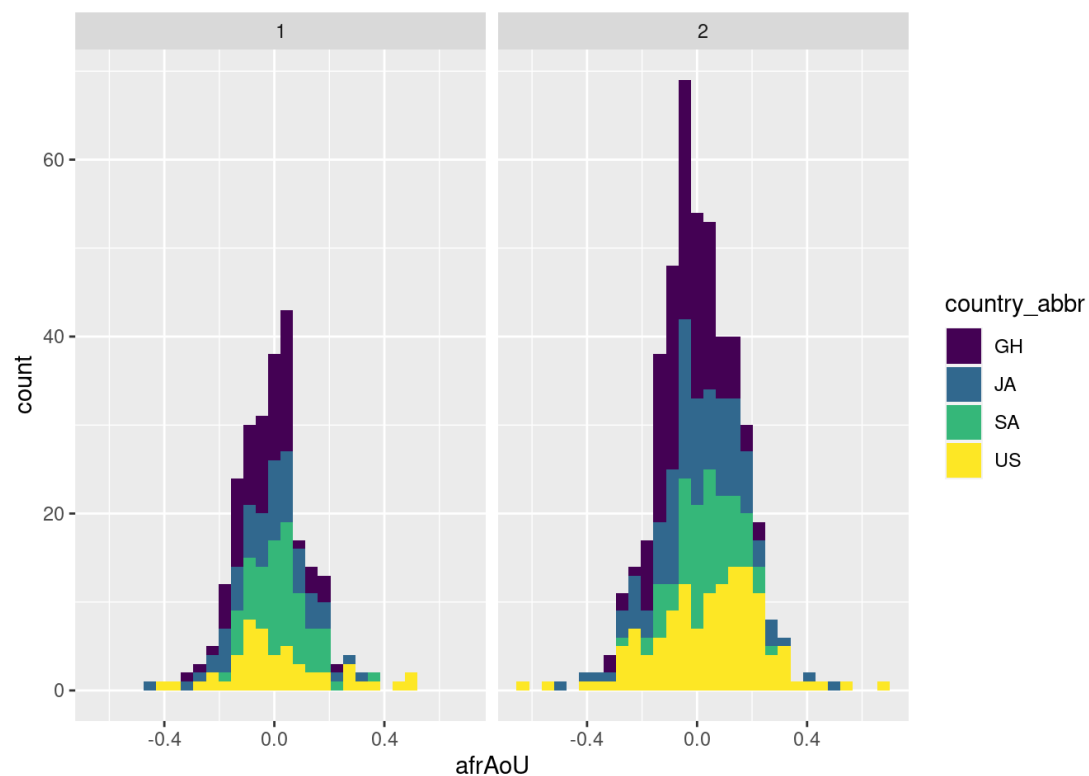


```
#age by site
ggplot(all_mets,aes(x=age,fill=country_abbr)) + geom_histogram() + scale_fill_viridis_d() + facet_wrap(~sex)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
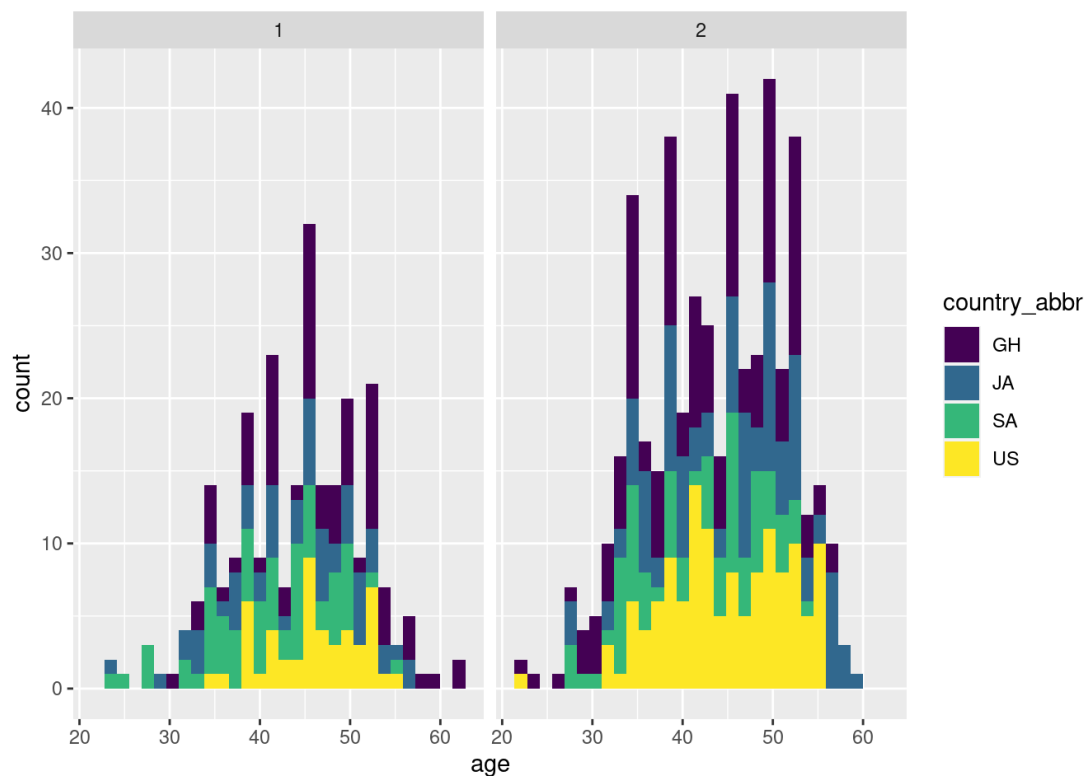
```
#scaled height by site
ggplot(all_mets,aes(x=scale(bmi),fill=country_abbr)) + geom_histogram() + scale_fill_viridis_d() + facet_wra
p(~sex)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_bin()`).
```
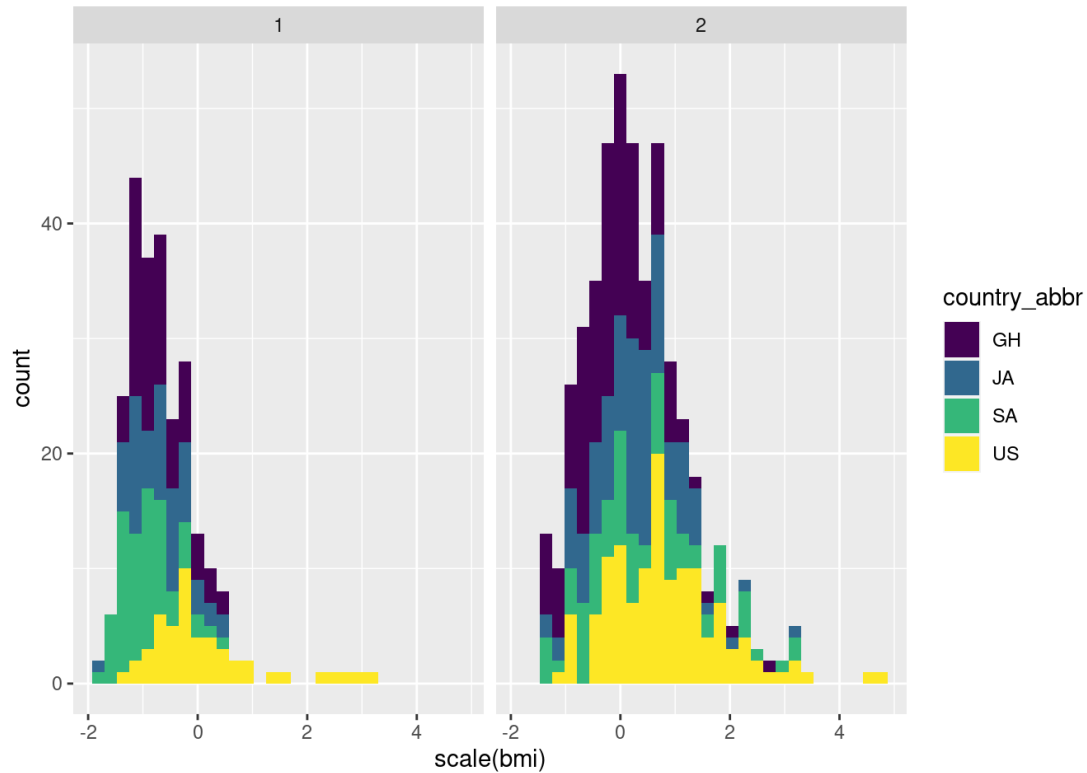


```
#allAoU PRS
ggplot(all_mets,aes(x=allAoU,fill=country_abbr)) + geom_histogram() + scale_fill_viridis_d() + facet_wrap(~s
ex)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
#afrAoU PRS
ggplot(all_mets,aes(x=afrAoU,fill=country_abbr)) + geom_histogram() + scale_fill_viridis_d() + facet_wrap(~s
ex)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
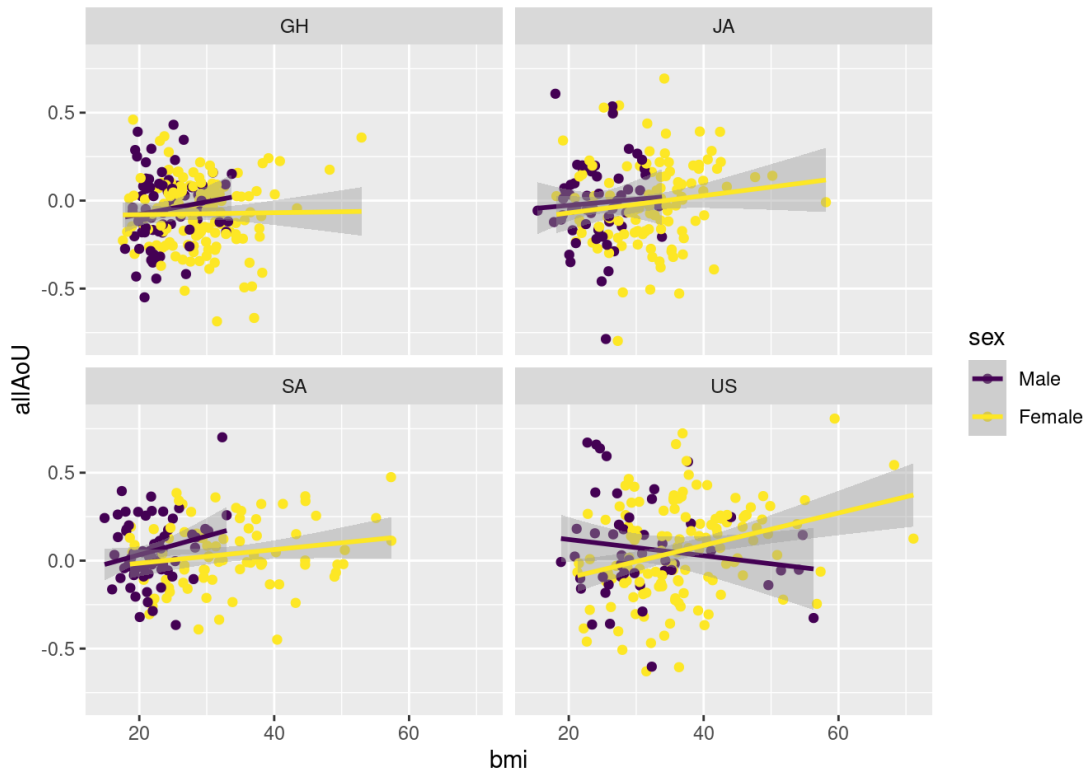


```
ggplot(all_mets,aes(x=bmi,y=allAoU,col=factor(sex,labels=c("Male","Female")))) + geom_point() + geom_smooth
(method='lm') + facet_wrap(~country_abbr) + guides(col=guide_legend(title="sex")) +  scale_color_viridis_d()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 6 rows containing missing values (`geom_point()`).
```
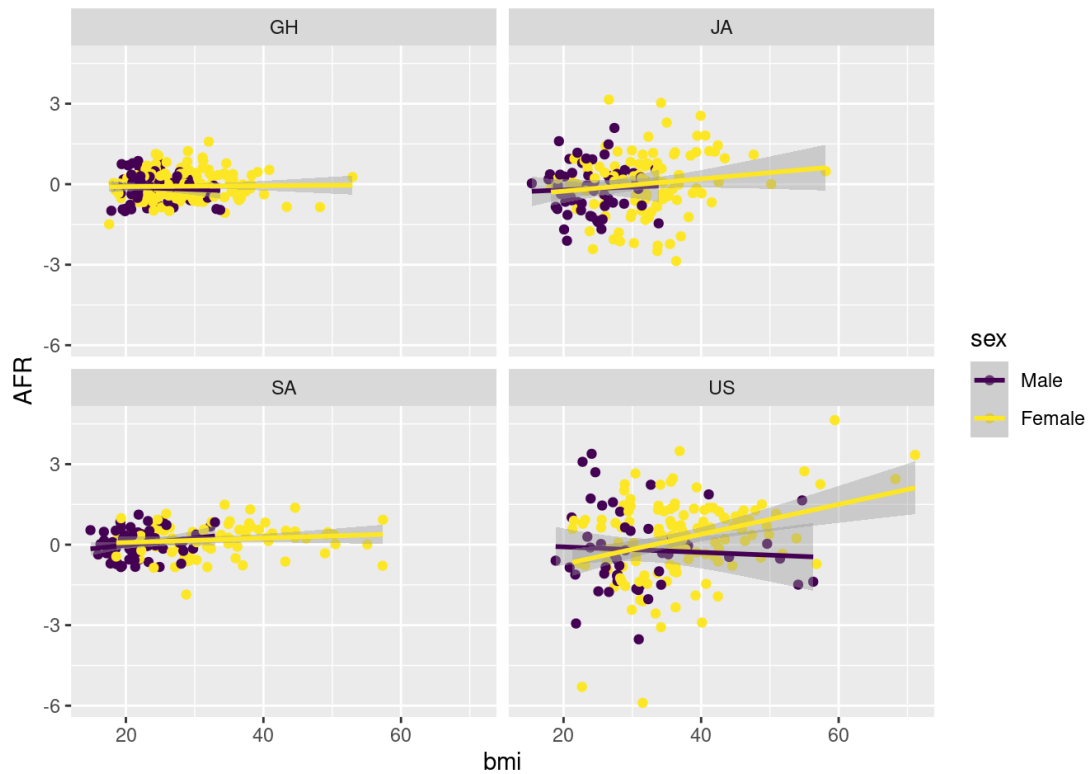


```
ggplot(all_mets,aes(x=bmi,y=AFR,col=factor(sex,labels=c("Male","Female")))) + geom_point() + geom_smooth(met
hod='lm') + facet_wrap(~country_abbr) + guides(col=guide_legend(title="sex"))+  scale_color_viridis_d()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_smooth()`).
## Removed 6 rows containing missing values (`geom_point()`).
```
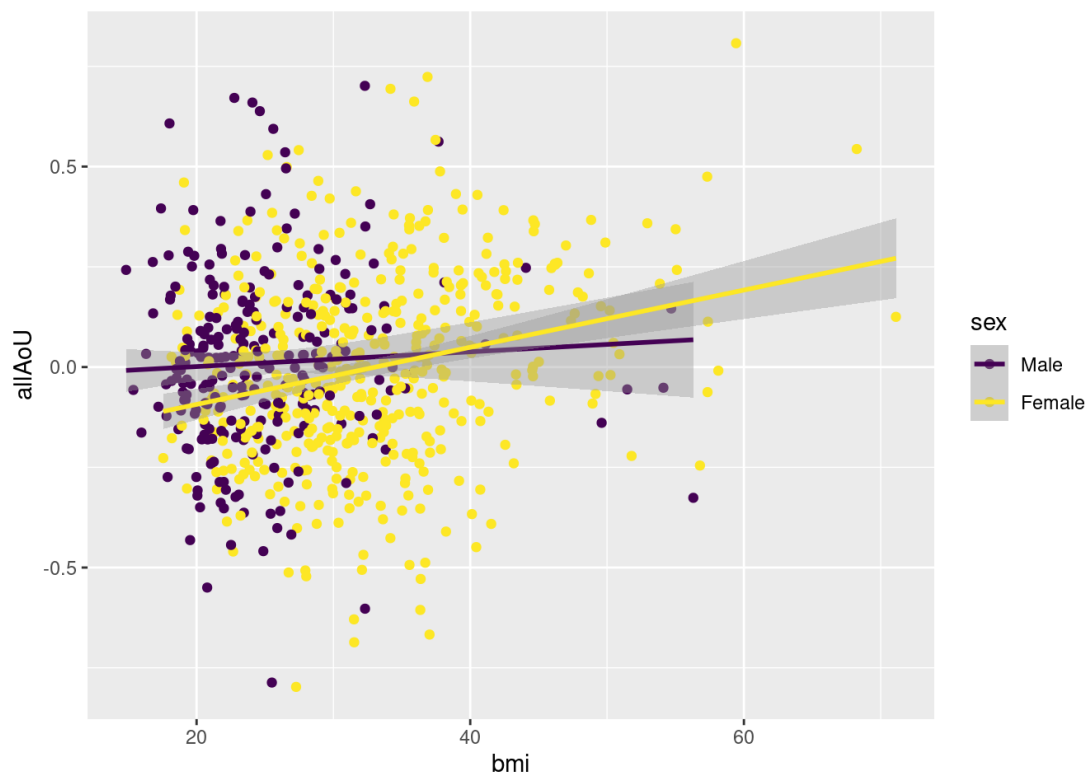
```
ggplot(all_mets,aes(x=bmi,y=allAoU,col=factor(sex,labels=c("Male","Female")))) + geom_point() + geom_smooth
(method='lm') + guides(col=guide_legend(title="sex"))+  scale_color_viridis_d()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_smooth()`).
## Removed 6 rows containing missing values (`geom_point()`).
```

```
ggplot(all_mets,aes(x=bmi,y=afrAoU,col=factor(sex,labels=c("Male","Female")))) + geom_point() + geom_smooth
(method='lm') + guides(col=guide_legend(title="sex"))+  scale_color_viridis_d()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 6 rows containing non-finite values (`stat_smooth()`).
## Removed 6 rows containing missing values (`geom_point()`).
```