

Supplemental material Rcode

Anna Kaal

2025-05-30

```
library("tidyverse")
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library("stringr")
library("DescTools")
library("rms")
```

```
## Loading required package: Hmisc
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:DescTools':
##
##      %nin%, Label, Mean, Quantile
##
## The following objects are masked from 'package:dplyr':
##
##      src, summarize
##
## The following objects are masked from 'package:base':
##
##      format.pval, units
```

```
library("VIM")
```

```
## Loading required package: colorspace
## Loading required package: grid
## VIM is ready to use.
##
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
```

```
##
## Attaching package: 'VIM'
##
## The following object is masked from 'package:datasets':
##
##     sleep
```

```
library("mice")
```

```
##
## Attaching package: 'mice'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
library("Matrix")
```

```
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
library("readr")
library("CalibrationCurves")
```

```
##
## Attaching package: 'CalibrationCurves'
##
## The following object is masked from 'package:VIM':
##
##     testdata
```

```
library("glmnet")
```

```
## Loaded glmnet 4.1-8
```

```
library("tableone")
```

The database michigan_predict contains all patients with a known PCT value with single imputation of missing data

```
michigan_predict <- read.csv("michigan_predict.csv")
michigan_predict$X <- NULL
```

First we do a visual inspection with the european data to decide the right value for the 'database' parameter

```
datasetsspainNL <- read.csv("../Predicted/totaaldatabase.csv")
datasetsspainNL_compare <- datasetsspainNL[c(2,4:8,13,19,20,22)]
michigan_predict_compare <- michigan_predict[c(1:9)]
michigan_predict_compare$BFRw_dich <- ifelse(michigan_predict_compare$RRw <=22,0,1)
michigan_predict_compare_final <- michigan_predict_compare[-4]
michigan_predict_compare_final<-michigan_predict_compare_final %>%
  mutate(HFRw = PULSEw) %>%
  mutate(AGE = AGEw*10) %>%
  mutate(EMVw = GCSw)

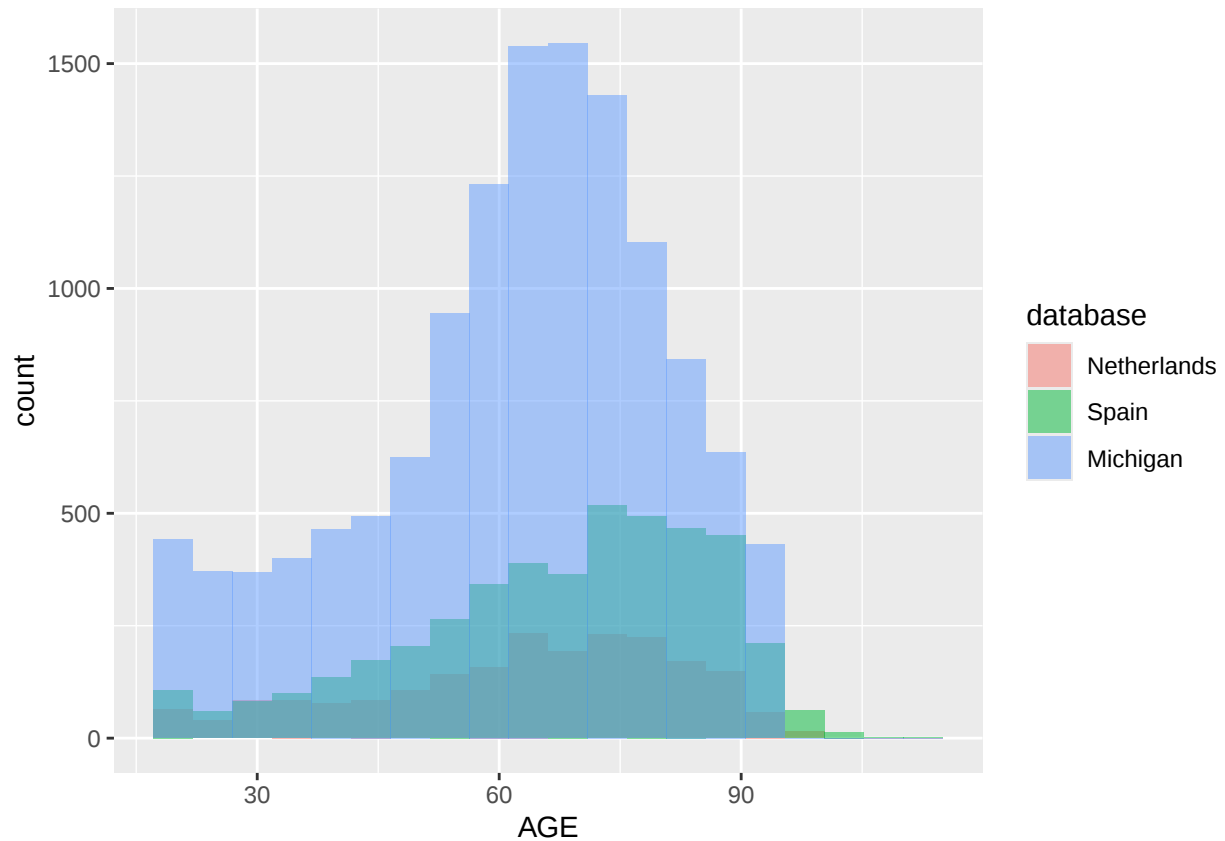
michigan_predict_compare_final$database <- 2
michigan_predict_compare_final2 <- michigan_predict_compare_final[-c(1,2,4)]
totaaldatabase <-rbind(datasetsspainNL_compare,michigan_predict_compare_final2)
table(totaaldatabase$database)
```

```
##
##      0      1      2
## 2111 4436 12864
```

```
totaaldatabase$database <- as.factor(totaaldatabase$database)

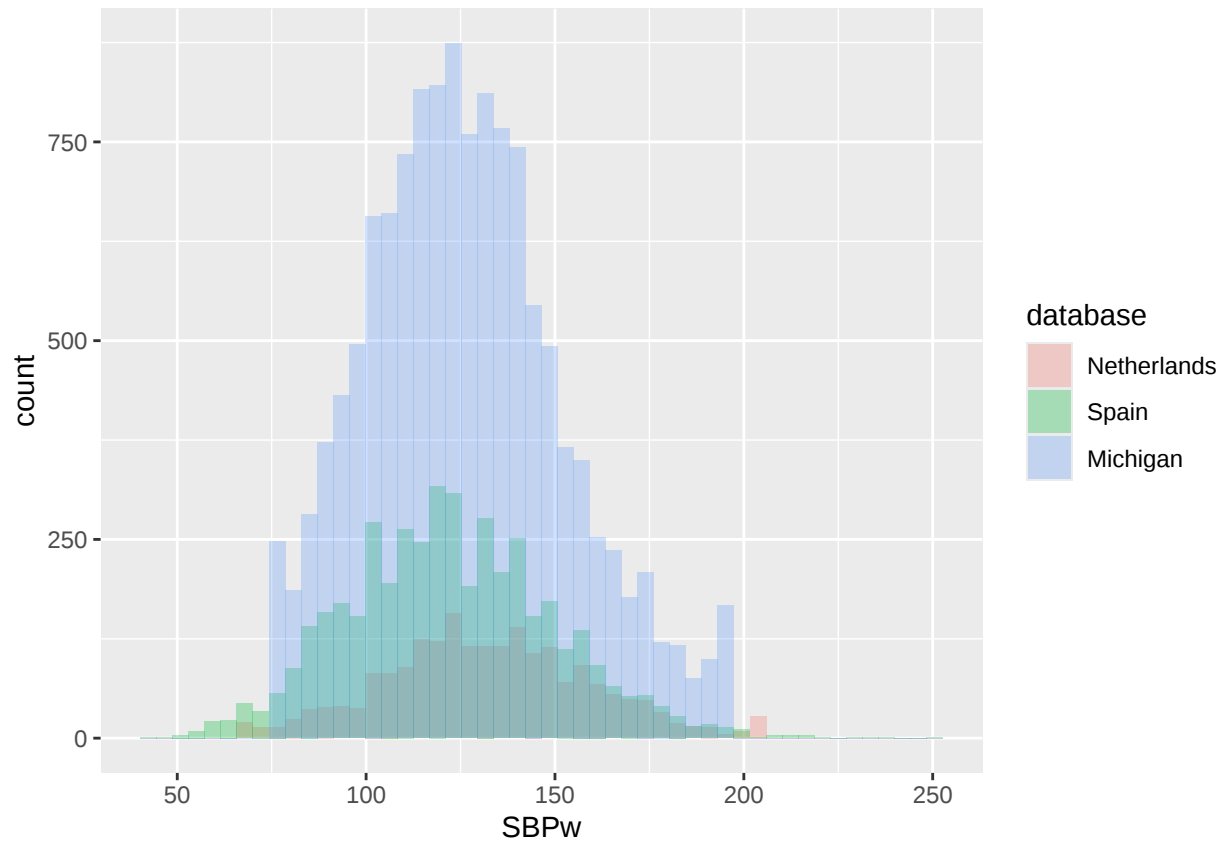
par(mfrow=c(5,5))

ggplot(totaaldatabase, aes(x = AGE, fill = database)) +
  geom_histogram(alpha = 0.5, position = "identity", bins = 20)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```



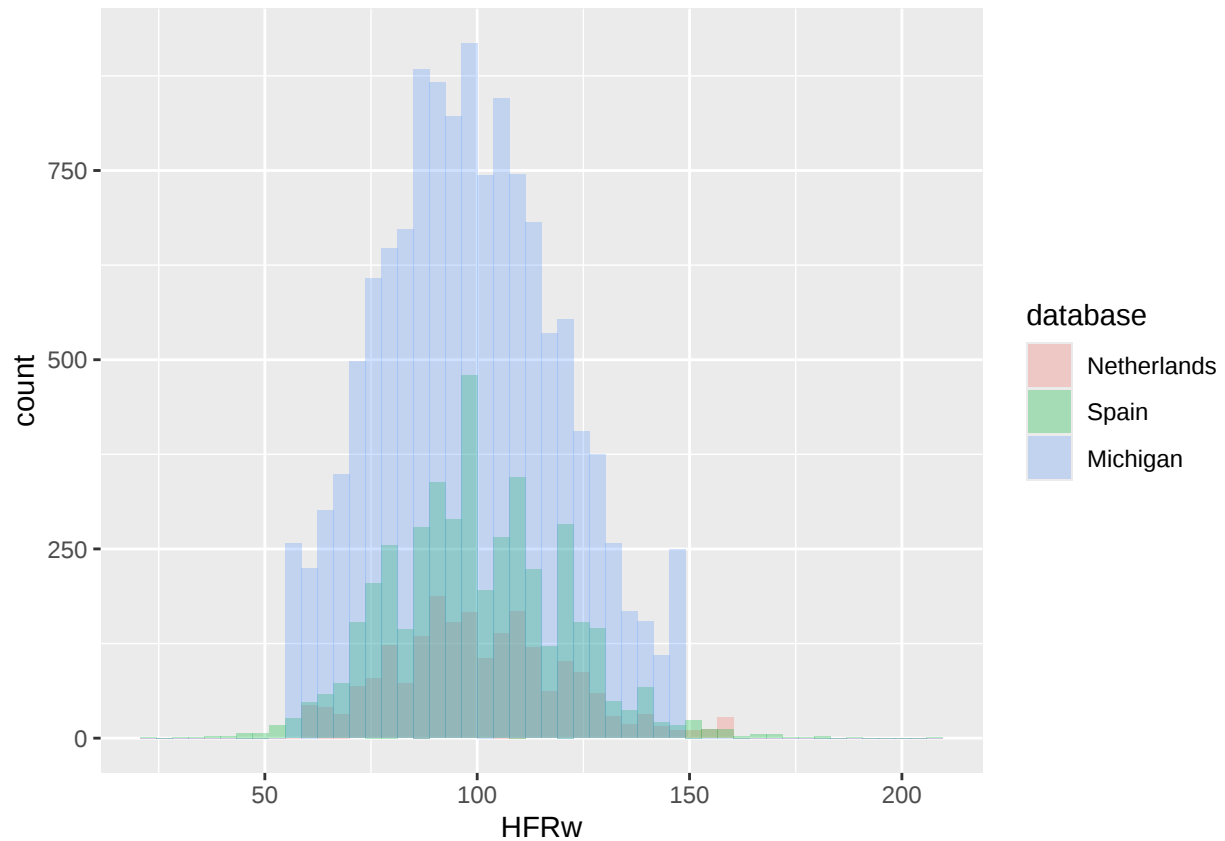
```
ggplot(totaaldatabase, aes(x = SBPw, fill = database)) +
  geom_histogram(alpha = 0.3, position = "identity", bins = 50)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```

```
## Warning: Removed 52 rows containing non-finite outside the scale range
## ('stat_bin()').
```



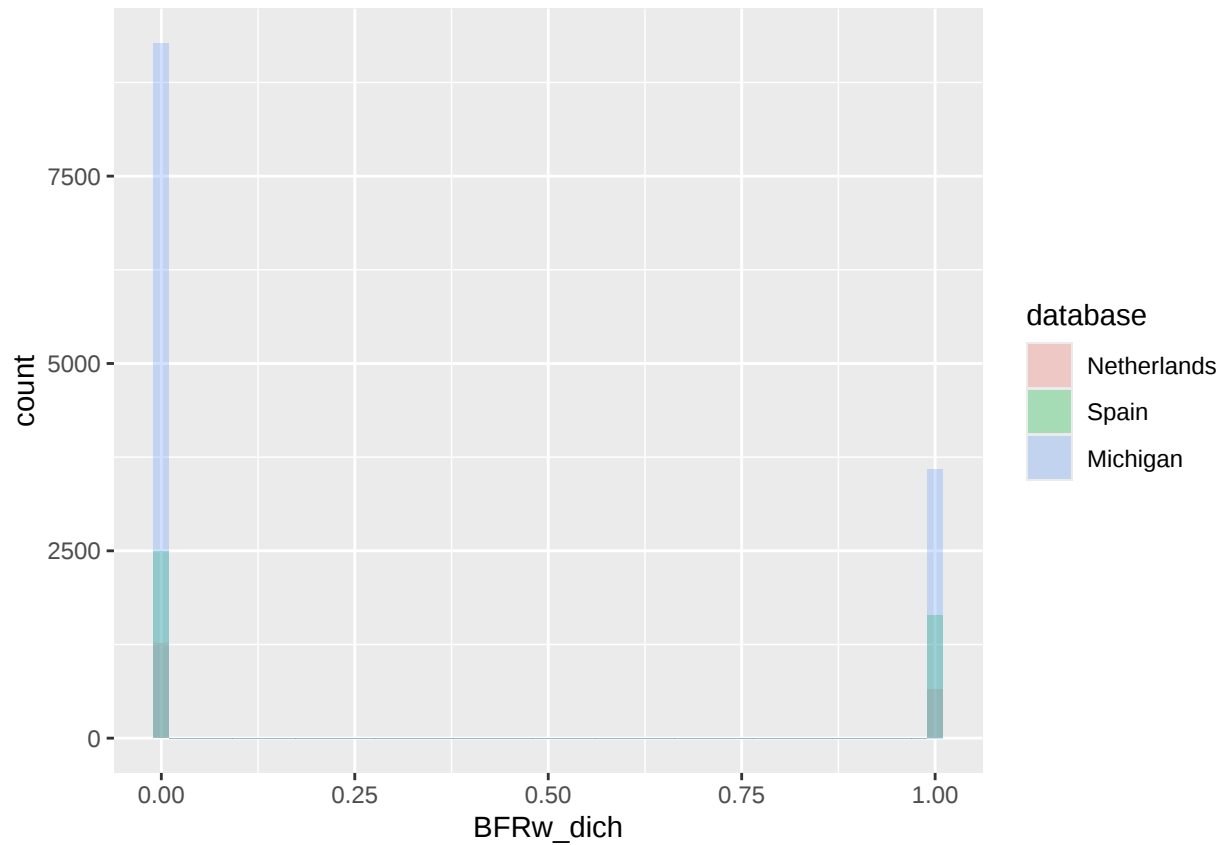
```
ggplot(totaaldatabase, aes(x = HFRw, fill = database)) +
  geom_histogram(alpha = 0.3, position = "identity", bins = 50)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```

```
## Warning: Removed 88 rows containing non-finite outside the scale range
## ('stat_bin()').
```



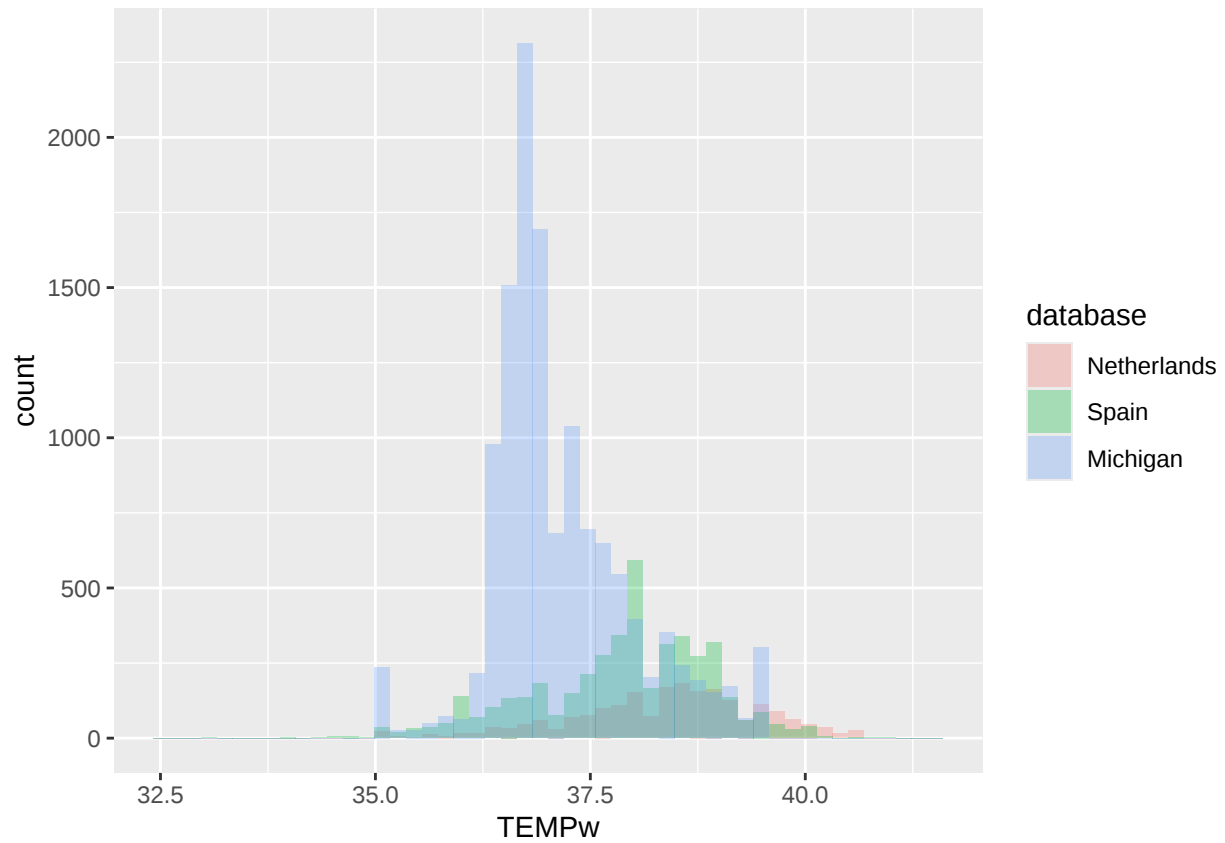
```
ggplot(totaaldatabase, aes(x = BFRw_dich, fill = database)) +
  geom_histogram(alpha = 0.3, position = "identity", bins = 50)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```

```
## Warning: Removed 504 rows containing non-finite outside the scale range
## ('stat_bin()').
```



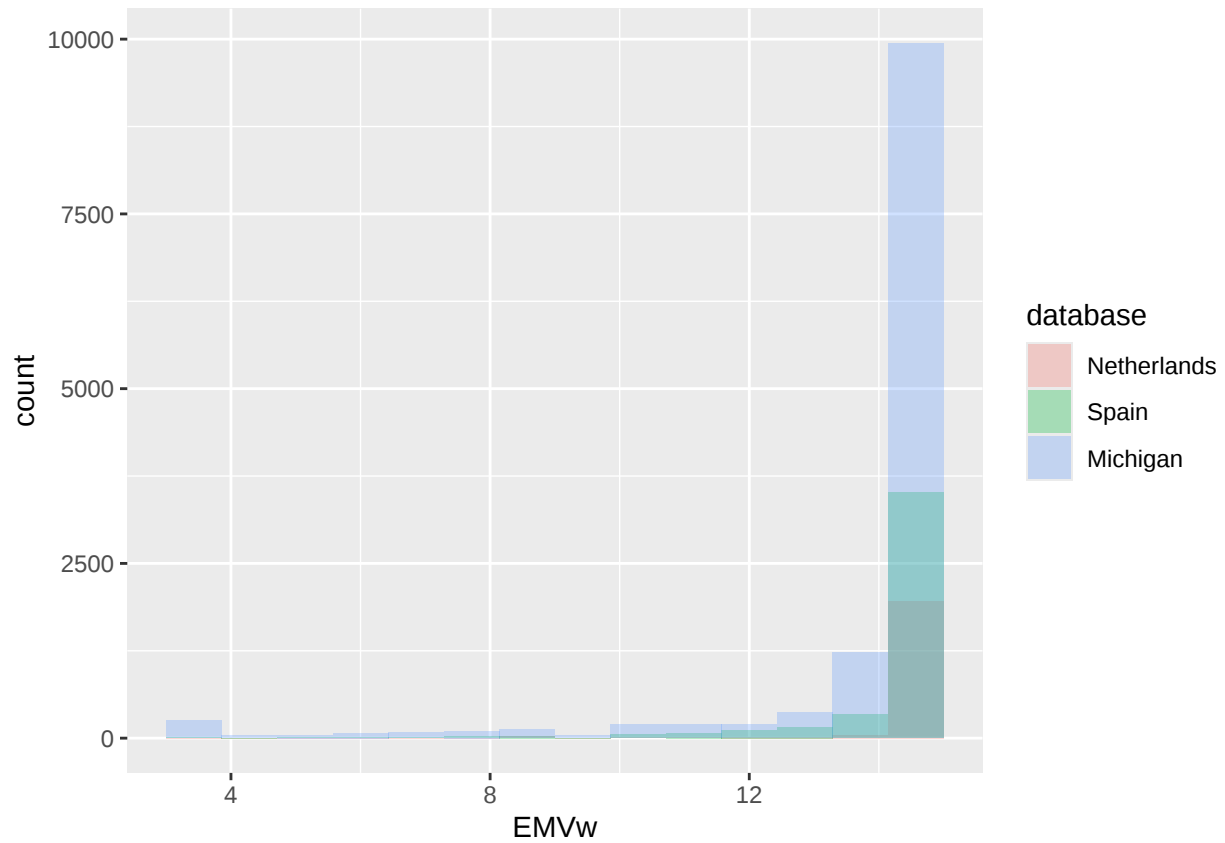
```
ggplot(totaaldatabase, aes(x = TEMPw, fill = database)) +
  geom_histogram(alpha = 0.3, position = "identity", bins = 50)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```

```
## Warning: Removed 17 rows containing non-finite outside the scale range
## ('stat_bin()').
```



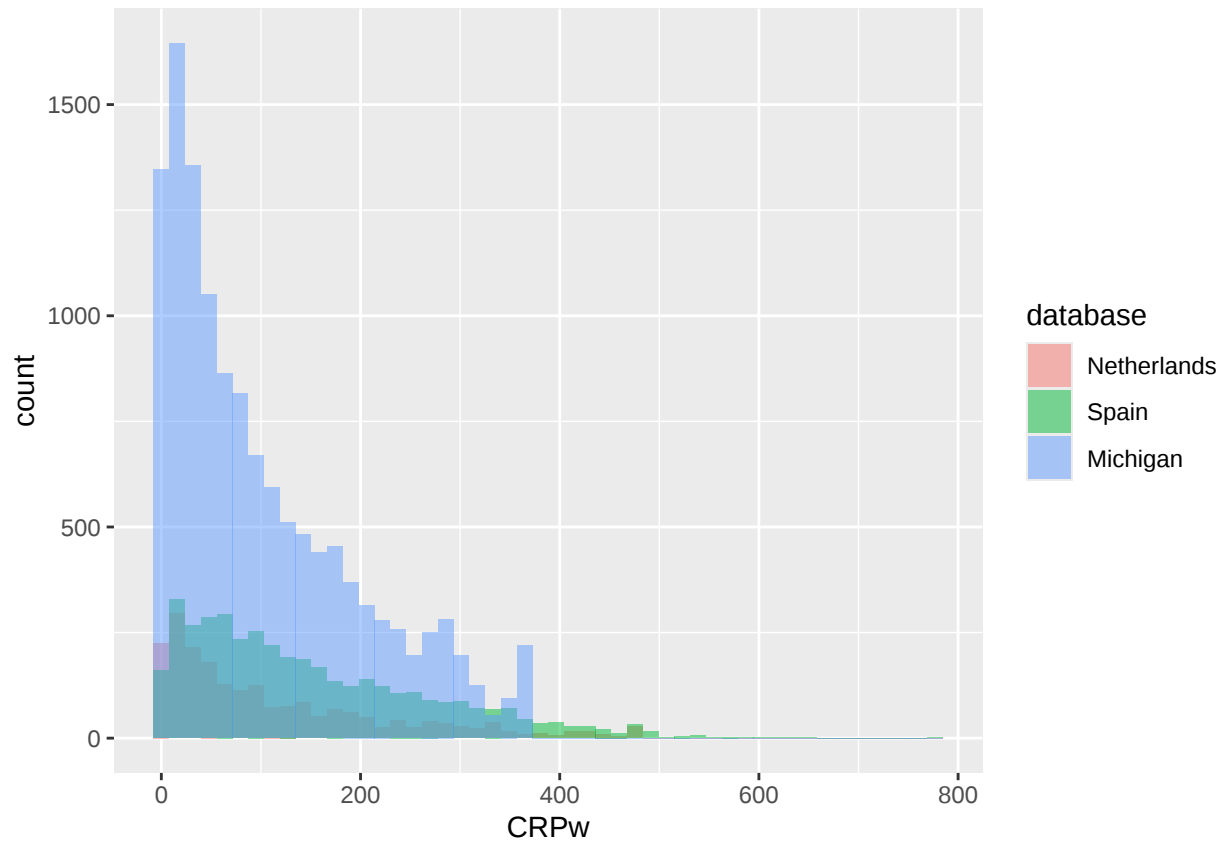
```
ggplot(totaaldatabase, aes(x = EMVw, fill = database)) +
  geom_histogram(alpha = 0.3, position = "identity", bins = 15)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```

```
## Warning: Removed 118 rows containing non-finite outside the scale range
## ('stat_bin()').
```



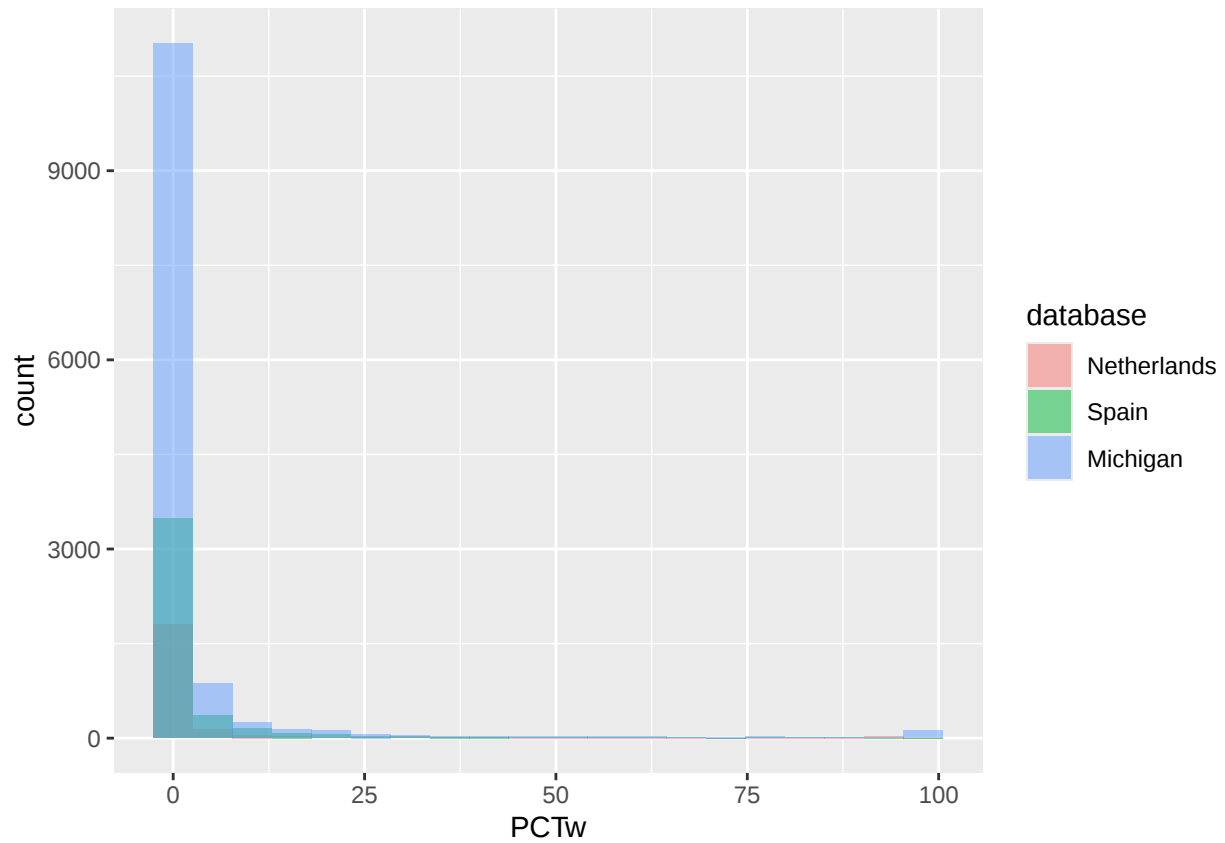
```
ggplot(totaaldatabase, aes(x = CRPw, fill = database)) +
  geom_histogram(alpha = 0.5, position = "identity", bins = 50)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```

```
## Warning: Removed 362 rows containing non-finite outside the scale range
## ('stat_bin()').
```

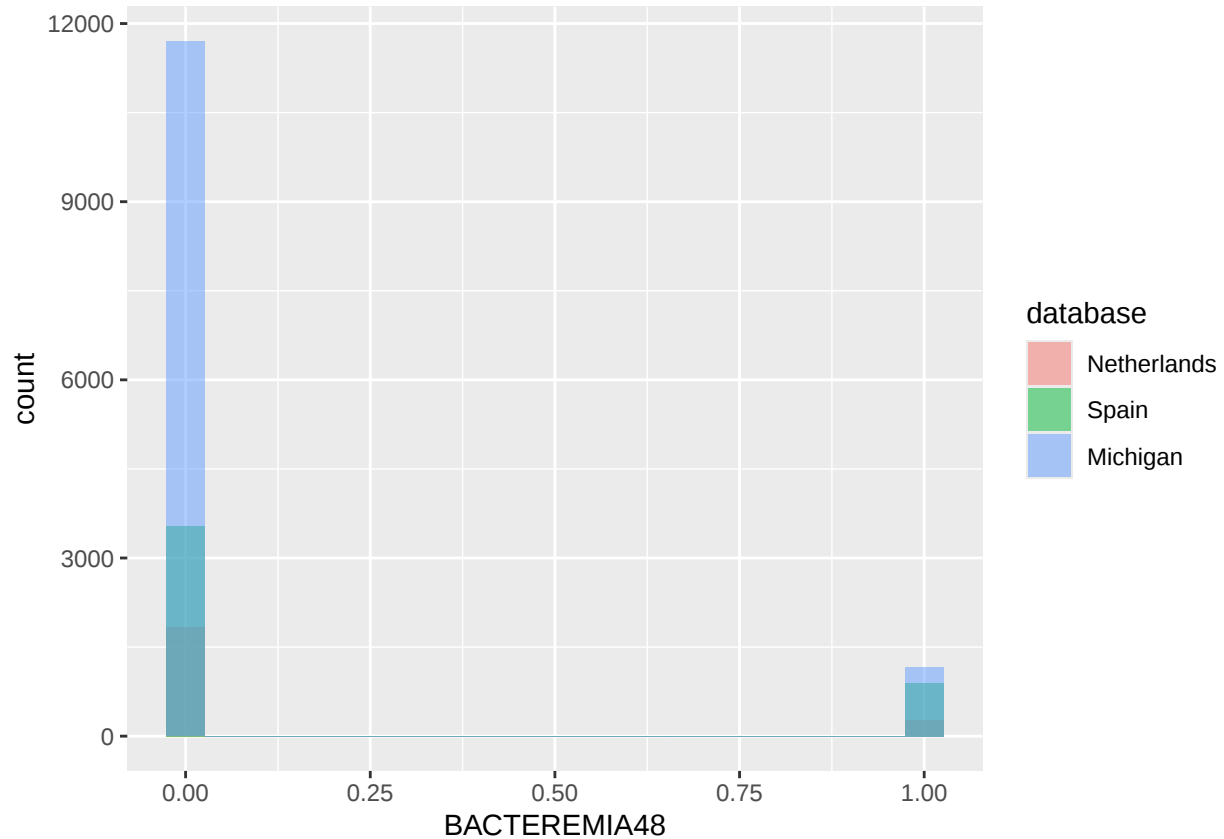


```
ggplot(totaaldatabase, aes(x = PCTw, fill = database)) +
  geom_histogram(alpha = 0.5, position = "identity", bins = 20)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```

```
## Warning: Removed 140 rows containing non-finite outside the scale range
## ('stat_bin()').
```



```
ggplot(totaaldatabase, aes(x = BACTEREMIA48, fill = database)) +
  geom_histogram(alpha = 0.5, position = "identity", bins = 20)+
  scale_fill_discrete(labels= c("Netherlands", "Spain", "Michigan"))
```



```
myvars2 <-c("AGE", "TEMPw", "HFRw", "SBPw", "EMVw", "BFRw_dich", "CRPw", "PCTw", "BACTEREMIA48", "database")
catVars2 <- c("BACTEREMIA48","database", "BFRw_dich")

baselinetable_compare <- CreateTableOne(vars=myvars2,data=totaaldatabase, factorVars = catVars2)

nonparametricvars2 <- c("AGE", "TEMPw", "HFRw", "SBPw", "EMVw", "CRPw", "PCTw")
print(baselinetable_compare, nonnormal = nonparametricvars2, formatOptions = list(big.mark = ","))
```

```
##
##
##      Overall
##      n
##      AGE (median [IQR])    66.00 [52.70, 77.00]
##      TEMPw (median [IQR])  37.20 [36.70, 38.10]
##      HFRw (median [IQR])   98.00 [84.00, 113.00]
##      SBPw (median [IQR])  124.33 [107.00, 143.00]
##      EMVw (median [IQR])   15.00 [15.00, 15.00]
##      BFRw_dich = 1 (%)     5879 (31.1)
##      CRPw (median [IQR])   84.00 [30.00, 175.00]
##      PCTw (median [IQR])    0.25 [0.10, 0.96]
##      BACTEREMIA48 = 1 (%)  2333 (12.0)
##      database (%)
##      0                      2111 (10.9)
##      1                      4436 (22.9)
##      2                      12864 (66.3)
```

```
# Per database
```

```
baselinetable_compare2 <- CreateTableOne(vars = myvars2, strata = "database", data = totaal database, fa
print(baselinetable_compare2, nonnormal = nonparametricvars2, formatOptions = list(big.mark = ","))
```

```
##          Stratified by database
##          0          1
##  n          2,111          4,436
##  AGE (median [IQR]) 66.00 [51.00, 78.00] 70.00 [56.00, 82.00]
##  TEMPw (median [IQR]) 38.50 [37.70, 39.10] 38.00 [37.20, 38.60]
##  HFRw (median [IQR]) 99.00 [87.00, 114.00] 100.00 [86.00, 114.00]
##  SBPw (median [IQR]) 131.00 [114.00, 150.00] 121.00 [103.00, 140.00]
##  EMVw (median [IQR]) 15.00 [15.00, 15.00] 15.00 [15.00, 15.00]
##  BFRw_dich = 1 (%) 649 ( 33.9) 1638 ( 39.7)
##  CRPw (median [IQR]) 74.00 [24.00, 170.50] 118.00 [54.00, 222.00]
##  PCTw (median [IQR]) 0.22 [0.09, 0.88] 0.32 [0.13, 1.34]
##  BACTEREMIA48 = 1 (%) 273 ( 12.9) 896 ( 20.2)
##  database (%)
##    0          2111 (100.0)          0 ( 0.0)
##    1           0 ( 0.0)        4436 (100.0)
##    2           0 ( 0.0)          0 ( 0.0)
##          Stratified by database
##          2          p          test
##  n          12,864
##  AGE (median [IQR]) 64.75 [51.71, 75.14] <0.001 nonnorm
##  TEMPw (median [IQR]) 36.90 [36.70, 37.60] <0.001 nonnorm
##  HFRw (median [IQR]) 98.00 [83.00, 113.00] <0.001 nonnorm
##  SBPw (median [IQR]) 124.50 [108.00, 142.00] <0.001 nonnorm
##  EMVw (median [IQR]) 15.00 [14.50, 15.00] <0.001 nonnorm
##  BFRw_dich = 1 (%) 3592 ( 27.9) <0.001
##  CRPw (median [IQR]) 75.00 [25.00, 162.00] <0.001 nonnorm
##  PCTw (median [IQR]) 0.23 [0.09, 0.87] <0.001 nonnorm
##  BACTEREMIA48 = 1 (%) 1164 ( 9.0) <0.001
##  database (%) <0.001
##    0           0 ( 0.0)
##    1           0 ( 0.0)
##    2        12864 (100.0)
```

More similar to dutch dataset so set database to 0

```
michigan_predict$database <- 0
```

import prediction models

```
PCTonly_model_update <- readRDS("../Predicted/model.PCTonly.update.rds")
basic0_model_update <- readRDS("../Predicted/model.basic0.update.rds")
basicCRP_model_update <- readRDS("../Predicted/model.basicCRP.update.rds")
basicPCT_model_update <- readRDS("../Predicted/model.basicPCT.update.rds")
basicCRP_PCT_model_update <- readRDS("../Predicted/model.basicCRP_PCT.update.rds")
```

```
nrow(michigan_predict)
```

```
## [1] 12864
```

```
table(michigan_predict$BACTEREMIA48, useNA="ifany")
```

```
##  
##      0      1  
## 11700  1164
```

Predictions for model with PCT

```
fit_ext_michigan_basicPCT <- glm(BACTEREMIA48~(ifelse(AGEw>7.5,7.5,AGEw))+(ifelse((SBPw/20)<6,(6-(SBPw/20)),6)),  
predict_michigan_basicPCT <- predict(basicPCT_model_update, newx = fit_ext_michigan_basicPCT$x, type = "cstat")  
Cstat(predict_michigan_basicPCT, michigan_predict$BACTEREMIA48)
```

```
## [1] 0.8034684
```

```
#Calibration in the large  
mean(predict_michigan_basicPCT)
```

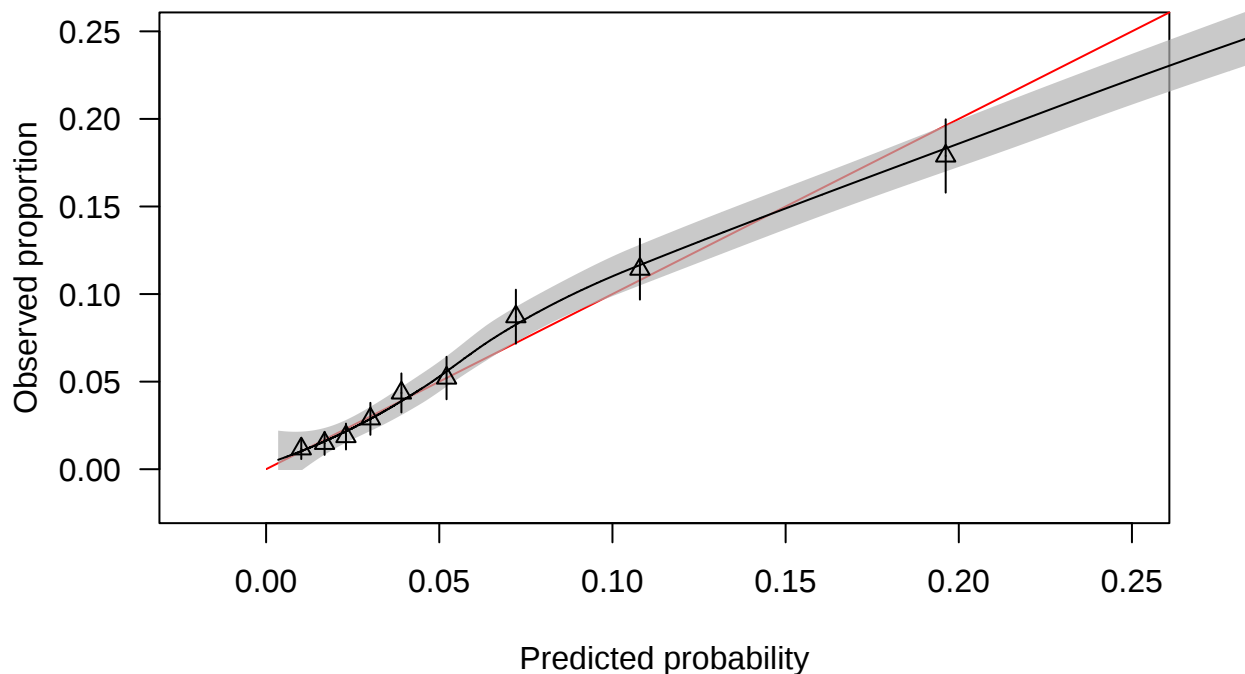
```
## [1] 0.1042852
```

```
mean(michigan_predict$BACTEREMIA48)
```

```
## [1] 0.09048507
```

```
#Calibration  
val.prob.ci.2(predict_michigan_basicPCT, michigan_predict$BACTEREMIA48, xlim = c(-0.02, 0.25), ylim = c(0.0, 1.0))
```

```
## Warning in val.prob.ci.2(predict_michigan_basicPCT,  
## michigan_predict$BACTEREMIA48, : Number of observations > 5000, RCS is  
## recommended. Will perform a preliminary fit to see if any errors occur.
```



```
## Call:
## val.prob.ci.2(p = predict_michigan_basicPCT, y = michigan_predict$BACTEREMIA48,
##   smooth = "loess", xlim = c(-0.02, 0.25), ylim = c(-0.02,
##     0.25), g = 10)
##
## A 95% confidence interval is given for the calibration intercept, calibration slope and c-statistic.
##
##           Dxy          C (ROC)          R2          D          D:Chi-sq
## 6.069367e-01 8.034684e-01 2.215125e-01 1.062056e-01 1.367229e+03
##           D:p          U          U:Chi-sq          U:p          Q
## 0.000000e+00 8.596784e-03 1.125890e+02 0.000000e+00 9.760886e-02
##           Brier      Intercept      Slope      Emax      Brier scaled
## 7.337495e-02 -2.057497e-01 7.944420e-01 1.671211e-01 1.084186e-01
##           Eavg          ECI
## 1.637137e-02 2.054716e-01
```

```
observations <- michigan_predict$BACTEREMIA48
drempelwaarden <- c(0.01,0.02, 0.025, 0.03, 0.04, 0.05,0.06, 0.07,0.08, 0.09, 0.10)
Threshold <- c()
sens <- c()
spec <- c()
NPV <- c()
PPV <- c()
Missed <- c()
Saved <-c()
```

```

SavedBC<- c()
Reduction <-c()

for (i in 1:11){
drempelwaarde <- drempelwaarden[i]
predicted_values<-ifelse(predict_michigan_basicPCT>drempelwaarde,1,0)
conf_matrix <- table(predicted_values, observations)
conf_matrix
Threshold[i] <- drempelwaarde
sens[i] <-(conf_matrix[2,2]/(conf_matrix[1,2]+conf_matrix[2,2]))*100 # wat is de kans dat een pt met een bacteriemi
spec[i] <-(conf_matrix[1,1]/(conf_matrix[1,1]+conf_matrix[2,1]))*100 # Wat is de kans dat een pt zonder bacteriemi
PPV[i] <- (conf_matrix[2,2]/(conf_matrix[2,1]+conf_matrix[2,2]))*100 # Wat is de kans op een bacteriemi
NPV[i] <- (conf_matrix[1,1]/(conf_matrix[1,1]+conf_matrix[1,2]))*100 # Wat is de kans op geen bacteriemi
Missed[i] <- conf_matrix[1,2]
Saved[i]<- conf_matrix[1,1]+conf_matrix[1,2]
SavedBC[i]<- Saved[i]*2
Reduction[i] <-((conf_matrix[1,1]+conf_matrix[1,2])/12864)*100
}

Michigan_Diagnostics.basicPCT <-as.data.frame(Threshold)
Michigan_Diagnostics.basicPCT$Sensitivity <- sens
Michigan_Diagnostics.basicPCT$Specificity <- spec
Michigan_Diagnostics.basicPCT$PPV <-PPV
Michigan_Diagnostics.basicPCT$NPV <-NPV
Michigan_Diagnostics.basicPCT$Missed <- Missed
Michigan_Diagnostics.basicPCT$MissedRelative <- Missed/1164*100
Michigan_Diagnostics.basicPCT$Saved <-Saved
Michigan_Diagnostics.basicPCT$SavedBC <-SavedBC
Michigan_Diagnostics.basicPCT$Reduction <- Reduction

Michigan_Diagnostics.basicPCT<-round(Michigan_Diagnostics.basicPCT,3)
Michigan_Diagnostics.basicPCT

```

##	Threshold	Sensitivity	Specificity	PPV	NPV	Missed	MissedRelative	Saved
## 1	0.010	99.570	4.795	9.424	99.117	5	0.430	566
## 2	0.020	97.079	21.880	11.003	98.689	34	2.921	2594
## 3	0.025	95.533	30.393	12.014	98.559	52	4.467	3608
## 4	0.030	93.814	37.897	13.065	98.402	72	6.186	4506
## 5	0.040	88.660	49.462	14.860	97.770	132	11.340	5919
## 6	0.050	85.739	57.496	16.714	97.592	166	14.261	6893
## 7	0.060	81.271	64.000	18.340	97.171	218	18.729	7706
## 8	0.070	77.663	68.393	19.644	96.853	260	22.337	8262
## 9	0.080	73.797	72.410	21.018	96.525	305	26.203	8777
## 10	0.090	70.017	75.479	22.123	96.198	349	29.983	9180
## 11	0.100	67.182	77.880	23.205	95.976	382	32.818	9494
##	SavedBC	Reduction						
## 1	1132	4.400						
## 2	5188	20.165						
## 3	7216	28.047						
## 4	9012	35.028						
## 5	11838	46.012						
## 6	13786	53.584						
## 7	15412	59.904						

```
## 8      16524      64.226
## 9      17554      68.229
## 10     18360      71.362
## 11     18988      73.803
```

Create calibration plot in PDF

```
# Step 1: Call the pdf command to start the plot
pdf(file = "Michigan_PCT.pdf", width = 6,height = 5)
```

```
# Step 2: Create the plot with R code
```

```
val.prob.ci.2(predict_michigan_basicPCT, michigan_predict$BACTEREMIA48, xlim = c(-0.02, 0.2), ylim = c(
```

```
## Warning in val.prob.ci.2(predict_michigan_basicPCT,
## michigan_predict$BACTEREMIA48, : Number of observations > 5000, RCS is
## recommended. Will perform a preliminary fit to see if any errors occur.
```

```
## Call:
```

```
## val.prob.ci.2(p = predict_michigan_basicPCT, y = michigan_predict$BACTEREMIA48,
##   smooth = "loess", xlab = "Predicted probability for basic model with PCT",
##   ylab = "Observed risk in the Michigan dataset", xlim = c(-0.02,
##     0.2), ylim = c(-0.02, 0.2), g = 10, legendloc = c(0.3,
##     1), statloc = c(0, 0.15))
##
```

```
## A 95% confidence interval is given for the calibration intercept, calibration slope and c-statistic
```

```
##
##           Dxy           C (ROC)           R2           D           D:Chi-sq
## 6.069367e-01 8.034684e-01 2.215125e-01 1.062056e-01 1.367229e+03
##           D:p           U           U:Chi-sq           U:p           Q
## 0.000000e+00 8.596784e-03 1.125890e+02 0.000000e+00 9.760886e-02
##           Brier      Intercept           Slope           Emax      Brier scaled
## 7.337495e-02 -2.057497e-01 7.944420e-01 1.671211e-01 1.084186e-01
##           Eavg           ECI
## 1.637137e-02 2.054716e-01
```

```
# Step 3: Run dev.off() to create the file!
dev.off()
```

```
## pdf
## 2
```