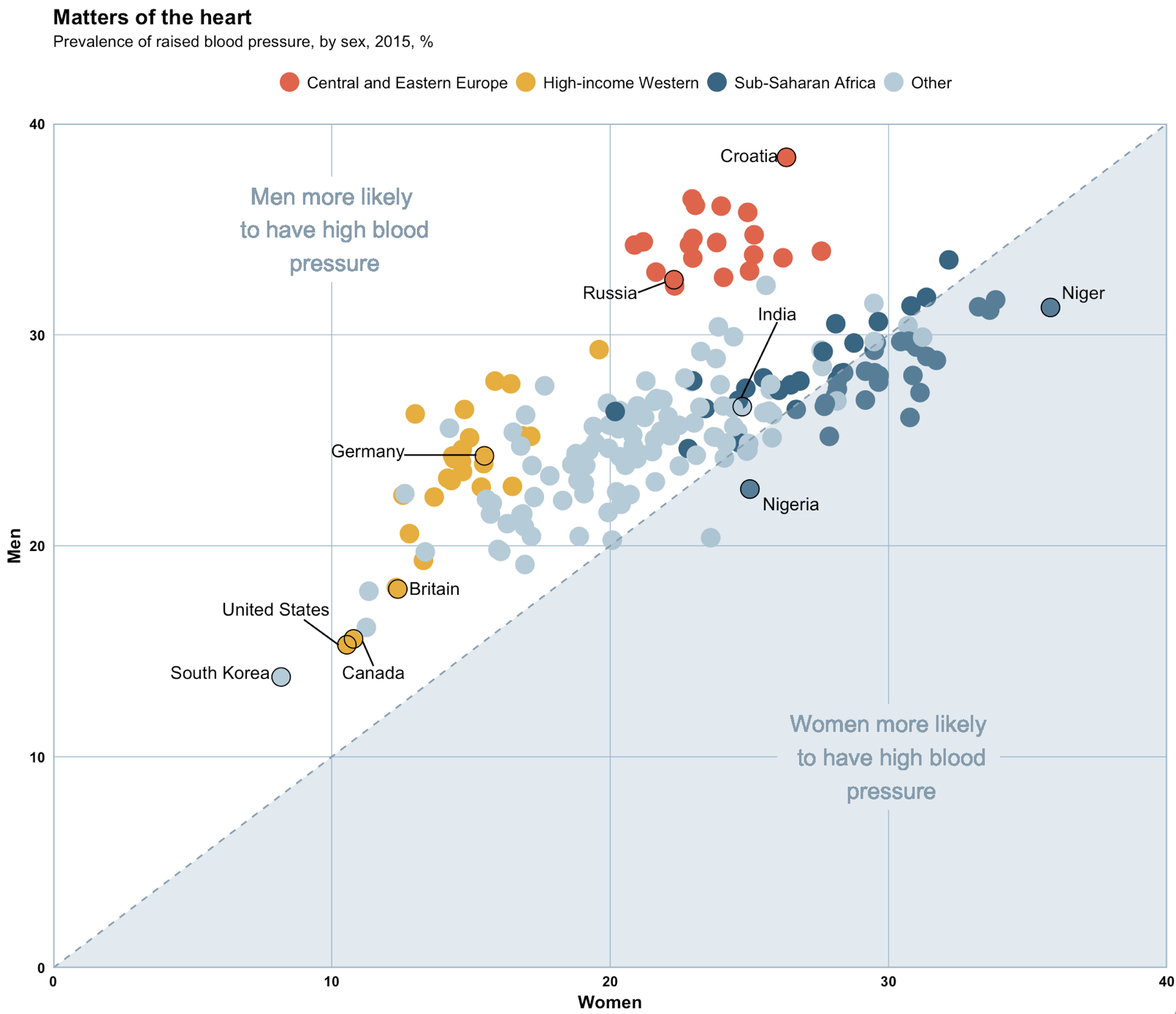
**(i) Display:**

****

**(ii) Origin of the data:**

This data comes from the NCD Risk Factor Collaboration (NCD RisC), which describes itself as “a network of health scientists around the world that provides rigorous and timely data on risk factors for non-communicable diseases (NCDs) for 200 countries and territories”.[[1]](#footnote-1) They pool data from many sources. The data used in this display is NCD RisC’s age-standardized, global data on blood pressure. It can be downloaded at <http://ncdrisc.org/data-downloads-blood-pressure.html> and was originally published in The Lancet at <https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(16)31919-5/fulltext>.

**(iii) Variables and Relationships in Context:**

There are three variables shown in the display for each of the individuals in this case, the 200 countries included in the NCD RisC data. Each country is represented as a data point which has an x-coordinate, a y-coordinate, and a color. The first variable, shown as the x-coordinate of each data point, is the percentage of women in a country who had raised blood pressure in 2015. The percentage of men in a country who had raised blood pressure in 2015 is graphed as the y-coordinate of each data point. Finally, the color of each data point displays the third variable, region, for which each country is categorized as “Central and Eastern Europe”, “High-income Western”, “Sub-Saharan Africa”, or “Other”.

So, the location of each data point displays the relationship between the 2015 prevalence of raised blood pressure for men and the same measurement for women. Those in the white upper-left section, to the left of the dotted line representing y=x, had a higher percentage of men with raised blood pressure than women. Likewise, the countries in the bluish lower-right section had a greater percentage of women than man with raised blood pressure. A majority of countries were in the first, white section of the display.

Interestingly, the relationship between a country’s region and the two blood pressure variables differs considerably by region. Countries from “Central and Eastern Europe” and “High-income Western” all had a larger percent of men with raised blood pressure than women, thus falling in the first section. But, the “High-income Western” countries had lower prevalence of raised blood pressure for both sexes than the “Central and Eastern Europe” countries. In the “Sub-Saharan Africa” region, countries have fairly even, fairly high percentages of men and women with raised blood pressure in 2015. But, they do fall on both sides of the y=x line, so approximately half have a slightly greater percentage of women with raised blood pressure than men. The “Other” countries exhibit the full range of raised blood pressure percentages by sex, but fall primarily in the first section, with a higher percentage of men with raised blood pressure.

**(iv) Code:**

library(mosaic)

#import dataset, data on blood pressure by sex, year, and country

NCD\_RisC\_Lancet\_2016\_BP\_age\_standardised\_countries <- read.csv("~/Downloads/NCD\_RisC\_Lancet\_2016\_BP\_age\_standardised\_countries.csv")

NCD <- NCD\_RisC\_Lancet\_2016\_BP\_age\_standardised\_countries

#prevalence of raised blood pressure by sex, year, and country

NCDp <- select(NCD, Country.Region.World,Sex,Year,Prevalence.of.raised.blood.pressure)

#prevalence of raised blood pressure by sex and country in 2015  
NCDp15 <- filter(NCDp,Year==2015)

#prevalence of raised blood pressure by country for men in 2015

men <- filter(NCDp15,Sex=="Men")

#prevalence of raised blood pressure by country for women in 2015

women <- filter(NCDp15,Sex=="Women")

#converts prevalence to percent from proportion

men <- mutate(men,mpercent=Prevalence.of.raised.blood.pressure\*100)

women <- mutate(women,wpercent=Prevalence.of.raised.blood.pressure\*100)

#put men and women in same dataframe in different columns

sbp <- data.frame(women$Country.Region.World,women$wpercent,men$mpercent)

colnames(sbp) <- c("country","wperc","mperc")

#new column for region, default “o” for region other

sbp <- mutate(sbp,region="o")

#value “a” in region for countries of sub-saharan africa

sbp$region[sbp$country=="Angola"] <- "a"

sbp$region[sbp$country=="Benin"] <- "a"

sbp$region[sbp$country=="Botswana"] <- "a"

sbp$region[sbp$country=="Burkina Faso"] <- "a"

sbp$region[sbp$country=="Burundi"] <- "a"

sbp$region[sbp$country=="Cabo Verde"] <- "a"

sbp$region[sbp$country=="Cameroon"] <- "a"

sbp$region[sbp$country=="Central African Republic"] <- "a"

sbp$region[sbp$country=="Chad"] <- "a"

sbp$region[sbp$country=="Comoros"] <- "a"

sbp$region[sbp$country=="Congo"] <- "a"

sbp$region[sbp$country=="Cote d'Ivoire"] <- "a"

sbp$region[sbp$country=="Djibouti"] <- "a"

sbp$region[sbp$country=="Equatorial Guinea"] <- "a"

sbp$region[sbp$country=="Eritrea"] <- "a"

sbp$region[sbp$country=="Ethiopia"] <- "a"

sbp$region[sbp$country=="Gabon"] <- "a"

sbp$region[sbp$country=="Gambia"] <- "a"

sbp$region[sbp$country=="Ghana"] <- "a"

sbp$region[sbp$country=="Guinea"] <- "a"

sbp$region[sbp$country=="Guinea Bissau"] <- "a"

sbp$region[sbp$country=="Kenya"] <- "a"

sbp$region[sbp$country=="Lesotho"] <- "a"

sbp$region[sbp$country=="Liberia"] <- "a"

sbp$region[sbp$country=="Madagascar"] <- "a"

sbp$region[sbp$country=="Malawi"] <- "a"

sbp$region[sbp$country=="Mali"] <- "a"

sbp$region[sbp$country=="Mauritania"] <- "a"

sbp$region[sbp$country=="Mauritius"] <- "a"

sbp$region[sbp$country=="Mozambique"] <- "a"

sbp$region[sbp$country=="Namibia"] <- "a"

sbp$region[sbp$country=="Niger"] <- "a"

sbp$region[sbp$country=="Nigeria"] <- "a"

sbp$region[sbp$country=="Rwanda"] <- "a"

sbp$region[sbp$country=="Sao Tome and Principe"] <- "a"

sbp$region[sbp$country=="Senegal"] <- "a"

sbp$region[sbp$country=="Seychelles"] <- "a"

sbp$region[sbp$country=="Sierra Leone"] <- "a"

sbp$region[sbp$country=="Somalia"] <- "a"

sbp$region[sbp$country=="South Africa"] <- "a"

sbp$region[sbp$country=="Sudan"] <- "a"

sbp$region[sbp$country=="Swaziland"] <- "a"

sbp$region[sbp$country=="Tanzania"] <- "a"

sbp$region[sbp$country=="Togo"] <- "a"

sbp$region[sbp$country=="Uganda"] <- "a"

sbp$region[sbp$country=="Zambia"] <- "a"

sbp$region[sbp$country=="Zimbabwe"] <- "a"

#value “e” in region for countries of central and eastern europe

sbp$region[sbp$country=="Albania"] <- "e"

sbp$region[sbp$country=="Belarus"] <- "e"

sbp$region[sbp$country=="Bosnia and Herzegovina"] <- "e"

sbp$region[sbp$country=="Bulgaria"] <- "e"

sbp$region[sbp$country=="Croatia"] <- "e"

sbp$region[sbp$country=="Czech Republic"] <- "e"

sbp$region[sbp$country=="Estonia"] <- "e"

sbp$region[sbp$country=="Hungary"] <- "e"

sbp$region[sbp$country=="Latvia"] <- "e"

sbp$region[sbp$country=="Lithuania"] <- "e"

sbp$region[sbp$country=="Macedonia (TFYR)"] <- "e"

sbp$region[sbp$country=="Moldova"] <- "e"

sbp$region[sbp$country=="Montenegro"] <- "e"

sbp$region[sbp$country=="Poland"] <- "e"

sbp$region[sbp$country=="Romania"] <- "e"

sbp$region[sbp$country=="Russian Federation"] <- "e"

sbp$region[sbp$country=="Serbia"] <- "e"

sbp$region[sbp$country=="Slovakia"] <- "e"

sbp$region[sbp$country=="Slovenia"] <- "e"

sbp$region[sbp$country=="Ukraine"] <- "e"

#value w in region for high-income western countries

sbp$region[sbp$country=="Andorra"] <- "w"

sbp$region[sbp$country=="Australia"] <- "w"

sbp$region[sbp$country=="Austria"] <- "w"

sbp$region[sbp$country=="Bahamas"] <- "w"

sbp$region[sbp$country=="Belgium"] <- "w"

sbp$region[sbp$country=="Canada"] <- "w"

sbp$region[sbp$country=="Cyprus"] <- "w"

sbp$region[sbp$country=="Denmark"] <- "w"

sbp$region[sbp$country=="Finland"] <- "w"

sbp$region[sbp$country=="France"] <- "w"

sbp$region[sbp$country=="Germany"] <- "w"

sbp$region[sbp$country=="Greece"] <- "w"

sbp$region[sbp$country=="Greenland"] <- "w"

sbp$region[sbp$country=="Iceland"] <- "w"

sbp$region[sbp$country=="Ireland"] <- "w"

sbp$region[sbp$country=="Israel"] <- "w"

sbp$region[sbp$country=="Italy"] <- "w"

sbp$region[sbp$country=="Liechtenstein"] <- "w"

sbp$region[sbp$country=="Luxembourg"] <- "w"

sbp$region[sbp$country=="Malta"] <- "w"

sbp$region[sbp$country=="Netherlands"] <- "w"

sbp$region[sbp$country=="New Zealand"] <- "w"

sbp$region[sbp$country=="Norway"] <- "w"

sbp$region[sbp$country=="Portugal"] <- "w"

sbp$region[sbp$country=="Spain"] <- "w"

sbp$region[sbp$country=="Sweden"] <- "w"

sbp$region[sbp$country=="Switzerland"] <- "w"

sbp$region[sbp$country=="United Kingdom"] <- "w"

sbp$region[sbp$country=="United States of America"] <- "w"

#expand values in region to full words, order them properly

sbp$region[sbp$region=="a"] <- "Sub-Saharan Africa"

sbp$region[sbp$region=="e"] <- "Central and Eastern Europe"

sbp$region[sbp$region=="w"] <- "High-income Western"

sbp$region[sbp$region=="o"] <- "Other"

sbp$region <- factor(sbp$region, levels=c("Central and Eastern Europe","High-income Western","Sub-Saharan Africa","Other"))

#initial scatterplot of data

A <- ggplot(sbp) + aes(x=wperc,y=mperc,color=region) + geom\_point(cex=5)

#add theme

A1 <- A + theme\_light()

#add labels to axes, title, subtitle

A2 <- A1 + labs(x="Women",y="Men",title="Matters of the heart",subtitle="Prevalence of raised blood pressure, by sex, 2015, %",color="")

#move legend

A3 <- A2 + theme(legend.position = "top")

#add line y=x

A4 <- A3 + geom\_abline(intercept=0,slope=1,linetype="dashed",color="#829daf")

#adjust graph scale

A5 <- A4 + scale\_x\_continuous(limits=c(0,40), expand = c(0,0)) + scale\_y\_continuous(limits=c(0,40), expand=c(0,0))

#change color scheme representing region

A6 <- A5 + scale\_color\_manual(values=c("tomato2","darkgoldenrod2","#2a6584","#b1cbd8"))

#blank out some grid lines

A7 <- A6 + annotate("rect",xmin=26,xmax=34,ymin=7.5,ymax=12.5,fill="white") + annotate("rect",xmin=7,xmax=13,ymin=32.5,ymax=37.5,fill="white")

#triangle points

x <- c(0,40,40)

y <- c(0,0,40)

tri <- data.frame(x,y)

#add shaded triangle to plot

A8 <- A7 + geom\_polygon(data=tri, aes(x=x,y=y),linetype="blank",color="#829daf",fill="#94b5cc",alpha=.3)

#add text explaining

A9 <- A8 + geom\_text(x=10,y=35,label="Men more likely\n to have high blood\n pressure",color="#829daf",size=5) + geom\_text(x=30,y=10,label="Women more likely\n to have high blood\n pressure",color="#829daf",size=5)

#theme changes

A10 <- A9 + theme(panel.grid.minor=element\_blank(),axis.ticks = element\_blank(),axis.text = element\_text(face="bold",color="black"),panel.border = element\_rect(color="#94b5cc",size=.7),panel.grid.major=element\_line(color="#94b5cc"),axis.title = element\_text(face="bold"),plot.title=element\_text(face="bold"),legend.text=element\_text(size=10))

#outlines of 10 key points

x <- c(8.183853,10.544834,10.78317,15.48522,12.373679,22.291300,26.33089,24.74478,25.017435,35.813458)

y <- c(13.78194,15.30833,15.58539,24.25953,17.94948,32.60418,38.40518,26.58928,22.68587,31.29110)

outlines <- data.frame(x,y)

A11 <- A10 + geom\_point(data=outlines,x=x,y=y,cex=5,pch=1,col="black")

#text labels for 10 key points

x <- c(6,8,11.5,13.7,11.3,26.5,37,25,20,26)

y <- c(14,17,14,18,24.5,22,32,38.5,32,31)

text <- c("South Korea","United States","Canada","Britain","Germany","Nigeria","Niger","Croatia","Russia","India")

label <- data.frame(x,y,text)

A12 <- A11 + geom\_text(data=label,x=x,y=y,label=text,color="black")

#line segments connecting labels to points for some key points

x <- c(9,11.5,12.6,21,26)

y <- c(16.5,14.3,24.3,32,30.6)

xend <- c(10.25,11.1,15.2,22,24.7)

yend <- c(15.3,15.5,24.3,32.5,27)

line <- data.frame(x,y,xend,yend)

A13 <- A12 + geom\_segment(data=line,x=x,y=y,xend=xend,yend=yend,color="black",lwd=.5)

#view the graph under “Zoom” so the text and line segments added are the right size

A13

1. . http://ncdrisc.org/about-us.html [↑](#footnote-ref-1)