Measurement and visualization, 1

Frank Edwards

9/24/2019

Data: Attitudes on Intimate Partner Violence

American Sociological Review
78(2) 240–265

DOI: 10.1177/0003122413480363
http://asr.sagepub.com

Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion

Rachael S. Pierotti[a]

Abstract

This study extends existing world society research on ideational diffusion by going beyond examinations of national policy change to investigate the spread of ideas among nonelite individuals. Specifically, I test whether recent trends in women's attitudes about intimate partner violence are converging toward global cultural scripts. Results suggest that global norms regarding violence against women are reaching citizens worldwide, including in some of the least privileged parts of the globe. During the first decade of the 2000s, women in 23 of the 26 countries studied became more likely to reject intimate partner violence. Structural socioeconomic or demographic changes, such as urbanization, rising educational attainment, increasing media access, and cohort replacement, fail to explain the majority of the observed trend. Rather, women of all ages and social locations became less likely to accept justifications for intimate partner violence. The near uniformity of the trend and speed of the change in attitudes about intimate partner violence suggest that global cultural diffusion has played an important role.

Reading in the data

```
ipv<-read_csv("./slides/data/dhs_ipv.csv")
```

The Data: USAID Demographic and Health Surveys

Name	Description
`beat_goesout`	Percentage of women in each country that think a husband is justified to beat his wife if she goes out without telling him.
`beat_burnfood`	Percentage of women in each country that think a husband is justified to beat his wife if she burns his food.
`no_media`	Percentage of women in each country that rarely encounter a newspaper, radio, or television.
`sec_school`	Percentage of women in each country with secondary or higher education.
`year`	Year of the survey
`region`	Region of the world
`country`	Country

Describing the data: country

```r
unique(ipv$country)
```

```
##  [1] "Albania"                  "Armenia"
##  [3] "Azerbaijan"               "Bangladesh"
##  [5] "Benin"                    "Bolivia"
##  [7] "Burkina Faso"             "Burundi"
##  [9] "Cambodia"                 "Cameroon"
## [11] "Chad"                     "Colombia"
## [13] "Comoros"                  "Congo (Brazzaville)"
## [15] "Congo Democratic Republic" "Cote d'Ivoire"
## [17] "Dominican Republic"       "Egypt"
## [19] "Eritrea"                  "Ethiopia"
## [21] "Gabon"                    "Gambia"
## [23] "Ghana"                    "Guinea"
## [25] "Guyana"                   "Haiti"
## [27] "Honduras"                 "India"
## [29] "Indonesia"                "Jordan"
## [31] "Kenya"                    "Kyrgyz Republic"
## [33] "Lesotho"                  "Liberia"
## [35] "Madagascar"               "Malawi"
## [37] "Maldives"                 "Mali"
## [39] "Mauritania"               "Moldova"
## [41] "Morocco"                  "Mozambique"
## [43] "Namibia"                  "Nepal"
## [45] "Nicaragua"                "Niger"
## [47] "Nigeria"                  "Pakistan"
## [49] "Peru"                     "Philippines"
## [51] "Rwanda"                   "Sao Tome and Principe"
## [53] "Senegal"                  "Sierra Leone"
```

Describing the data: country

```r
length(unique(ipv$country))
```

```
## [1] 65
```

```r
nrow(ipv)
```

```
## [1] 151
```

Describing the data: region

```
table(ipv$region)
```

```
##
##                          Asia                 Latin America
##                            24                            24
## Middle East and Central Asia           Sub-Saharan Africa
##                            19                            84
```

Describing the data: time

```
table(ipv$year)
```

```
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
##    1   14    5    5   11    8   17    8   10   10    9   12   13   14   11    3
```

Describing continuous measures: what do these measures show?

```
ipv %>%
  select(beat_burnfood,
         beat_goesout,
         sec_school,
         no_media) %>%
  summary()
```

```
##  beat_burnfood    beat_goesout     sec_school       no_media
##  Min.   : 0.10   Min.   : 0.30   Min.   : 3.10   Min.   : 0.80
##  1st Qu.: 4.50   1st Qu.:11.85   1st Qu.:10.18   1st Qu.:11.25
##  Median :11.85   Median :28.10   Median :22.40   Median :29.15
##  Mean   :15.04   Mean   :28.60   Mean   :24.40   Mean   :28.40
##  3rd Qu.:22.25   3rd Qu.:42.08   3rd Qu.:34.90   3rd Qu.:43.23
##  Max.   :64.50   Max.   :82.70   Max.   :74.60   Max.   :86.40
##  NA's   :31      NA's   :27      NA's   :3       NA's   :13
```

Missing data in R

Missing data as a single value

```
a<-NA
a+a
```

```
## [1] NA
```

Missing data as a single value

```r
a<-NA
a+a
```

```
## [1] NA
```

```r
a*2
```

```
## [1] NA
```

Missing data as a single value

```r
a<-NA
a+a
```

```
## [1] NA
```

```r
a*2
```

```
## [1] NA
```

```r
a==TRUE
```

```
## [1] NA
```

```r
is.na(a)
```

```
## [1] TRUE
```

```r
!(is.na(a))
```

```
## [1] FALSE
```

Missing data in vectors

```
my_cool_vector<-c(2, 3, NA, 4)
mean(my_cool_vector)
```

```
## [1] NA
```

```
min(my_cool_vector)
```

```
## [1] NA
```

Dealing with missing data in vectors

```r
mean(my_cool_vector, na.rm=TRUE)
```

```
## [1] 3
```

```r
min(my_cool_vector, na.rm=TRUE)
```

```
## [1] 2
```

```r
sd(my_cool_vector, na.rm=TRUE)
```

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

Dealing with missing data in practice

```
ipv %>%
  summarise(beat_burnfood_mn_bad =
              mean(beat_burnfood),
            beat_burnfood_mn_good =
              mean(beat_burnfood, na.rm = TRUE))
```

```
## # A tibble: 1 x 2
##   beat_burnfood_mn_bad beat_burnfood_mn_good
##                  <dbl>                 <dbl>
## 1                   NA                  15.0
```

Counting missing values

```
ipv %>%
  summarise(beat_burnfood_missing =
              sum(is.na(beat_burnfood))/n())
```

```
## # A tibble: 1 x 1
##   beat_burnfood_missing
##                   <dbl>
## 1                 0.205
```

```
table(is.na(ipv$beat_burnfood))
```

```
##
## FALSE  TRUE
##   120    31
```

Filtering out missing values

```
ipv_no_missing<-na.omit(ipv)

ipv_no_missing_onevar<-ipv %>%
  filter(!(is.na(beat_burnfood)))

nrow(ipv_no_missing)
```

```
## [1] 116
```

```
nrow(ipv_no_missing_onevar)
```

```
## [1] 120
```

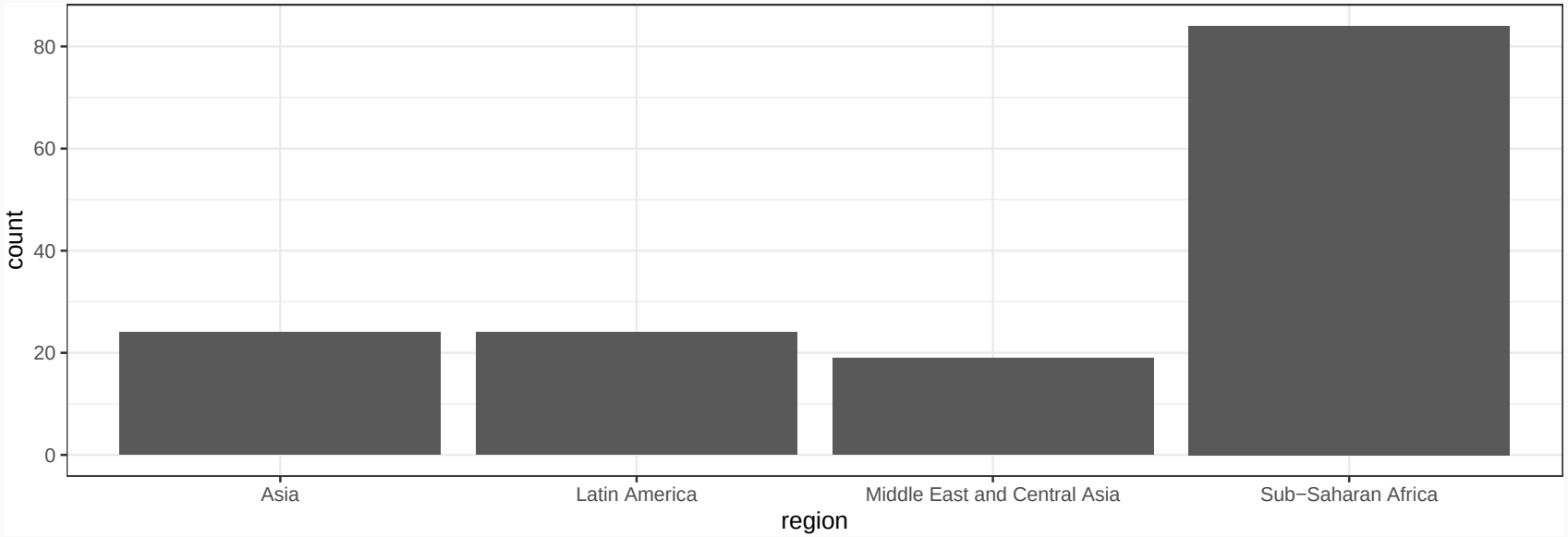
Visualizing the distribution of single variables (univariate visuals)

Visuals for categorical variables

Barplots

Show the count of rows in each value of a category

```
ggplot(ipv,
       aes(x = region)) +
  geom_bar()
```



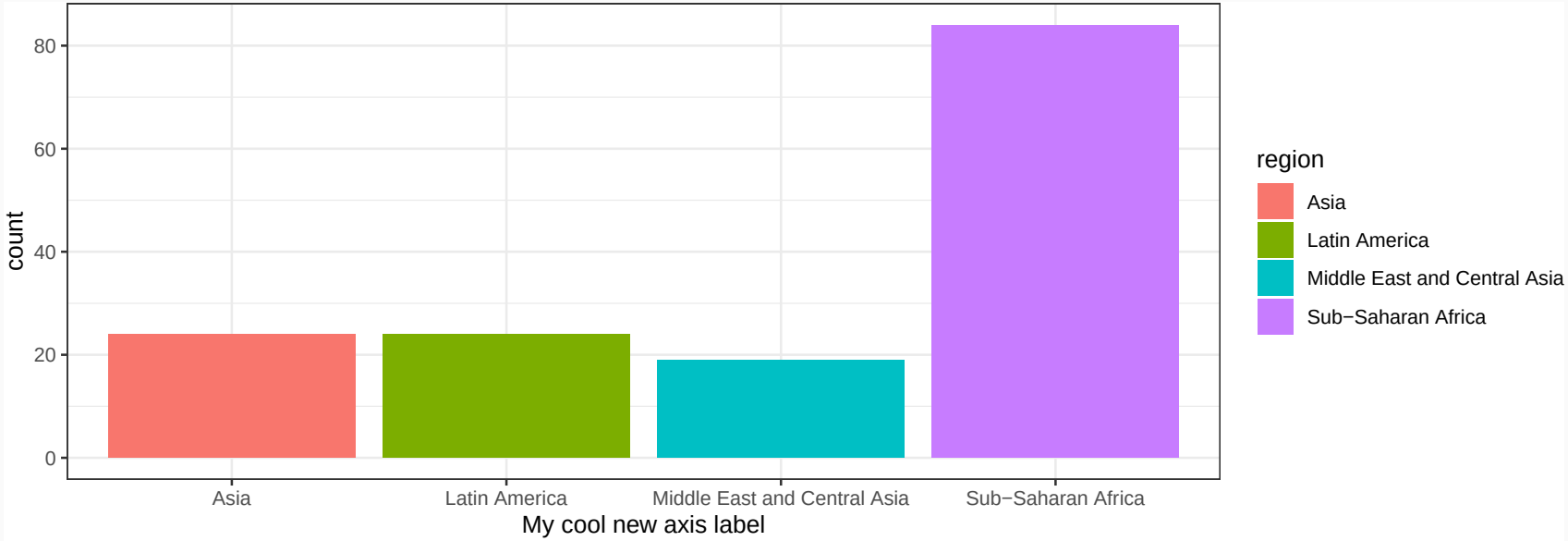
The anatomy of a basic ggplot() call

```
ggplot(ipv, ## object, usually a data.frame
       aes(x = region)) + ## aesthetic variables, generally x, y, color, etc
  geom_bar() ## a geom to plot the aesthetics

## Note that + in ggplot() works the same way as %>% in tidyverse:
## It strings together commands, evaluated in sequence
```

Adding to our call

```
ggplot(ipv,
       aes(x = region)) +
  geom_bar(aes(fill = region)) +
  xlab("My cool new axis label")
```

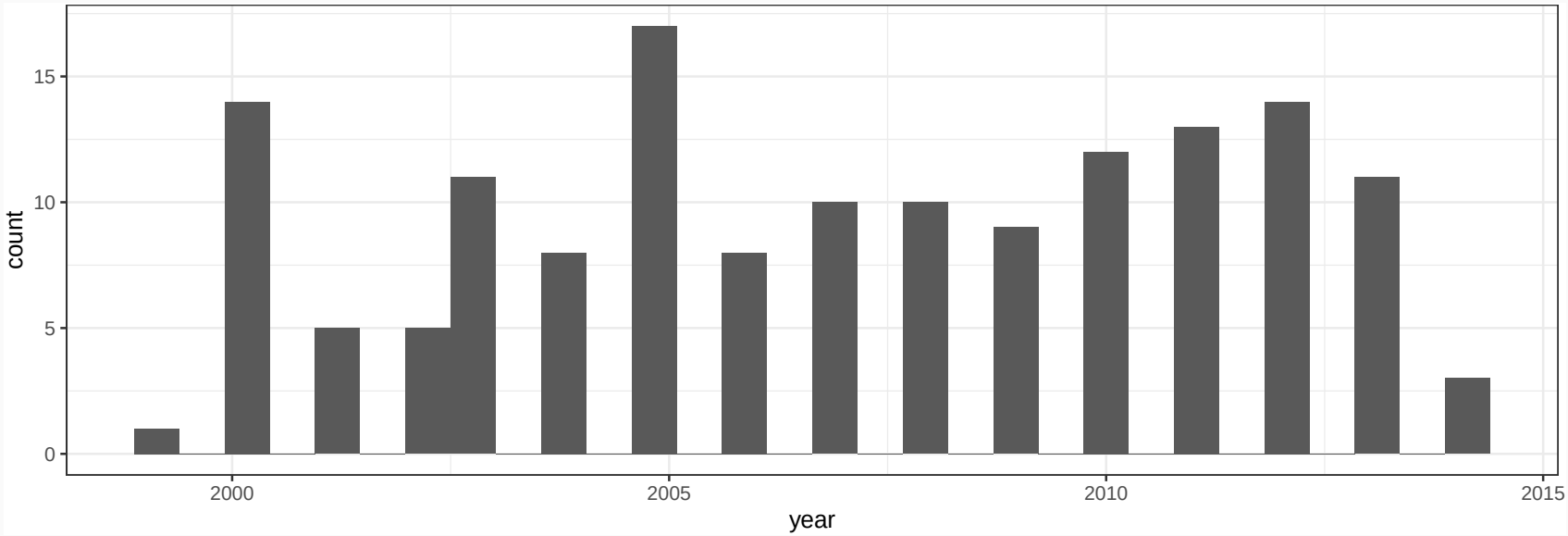


Visuals for continuous variables

Histograms

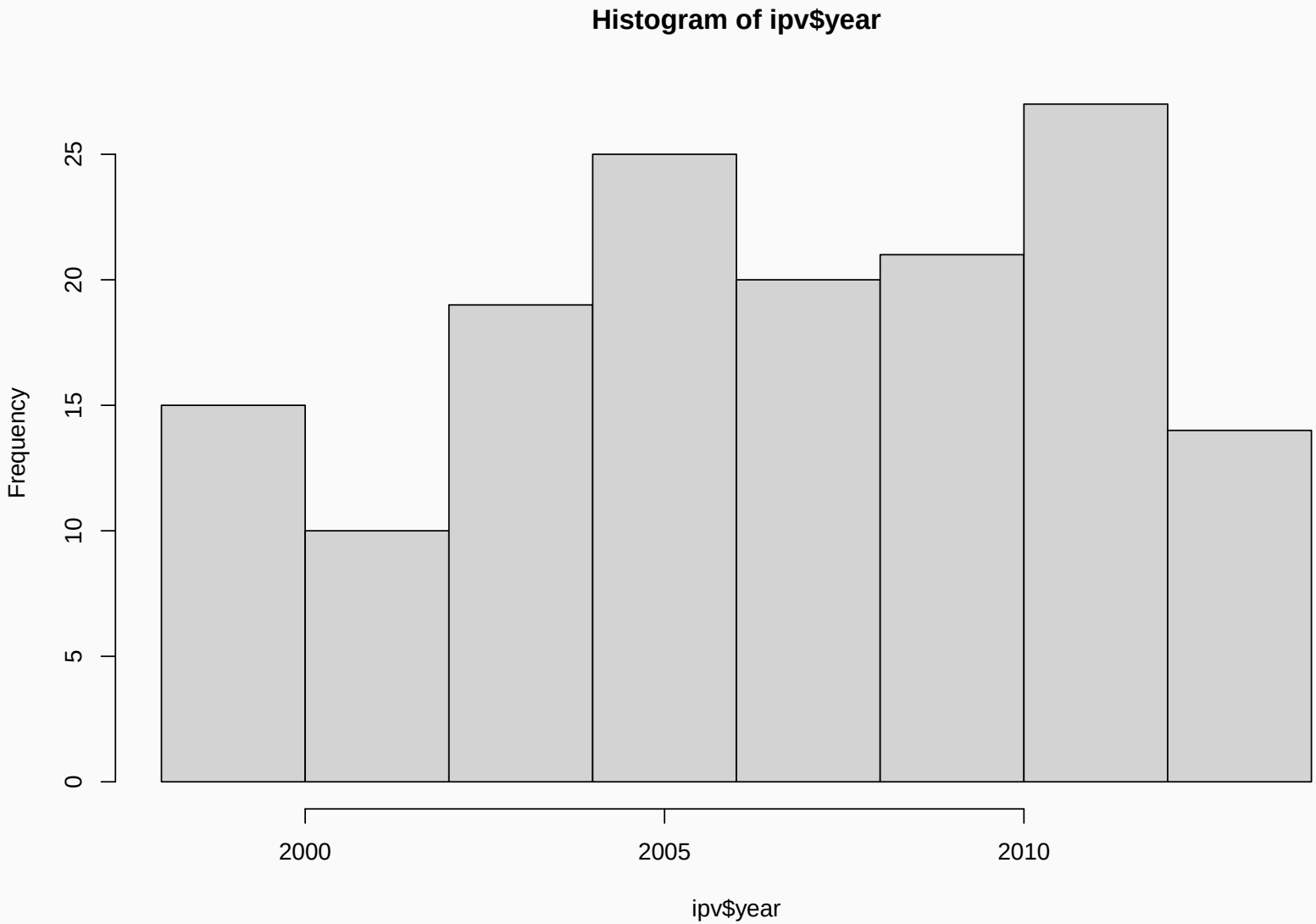
Histograms show the density of cases that fall within a given range

```
ggplot(ipv,
       aes(x = year)) +
  geom_histogram()
```



Histograms in base R

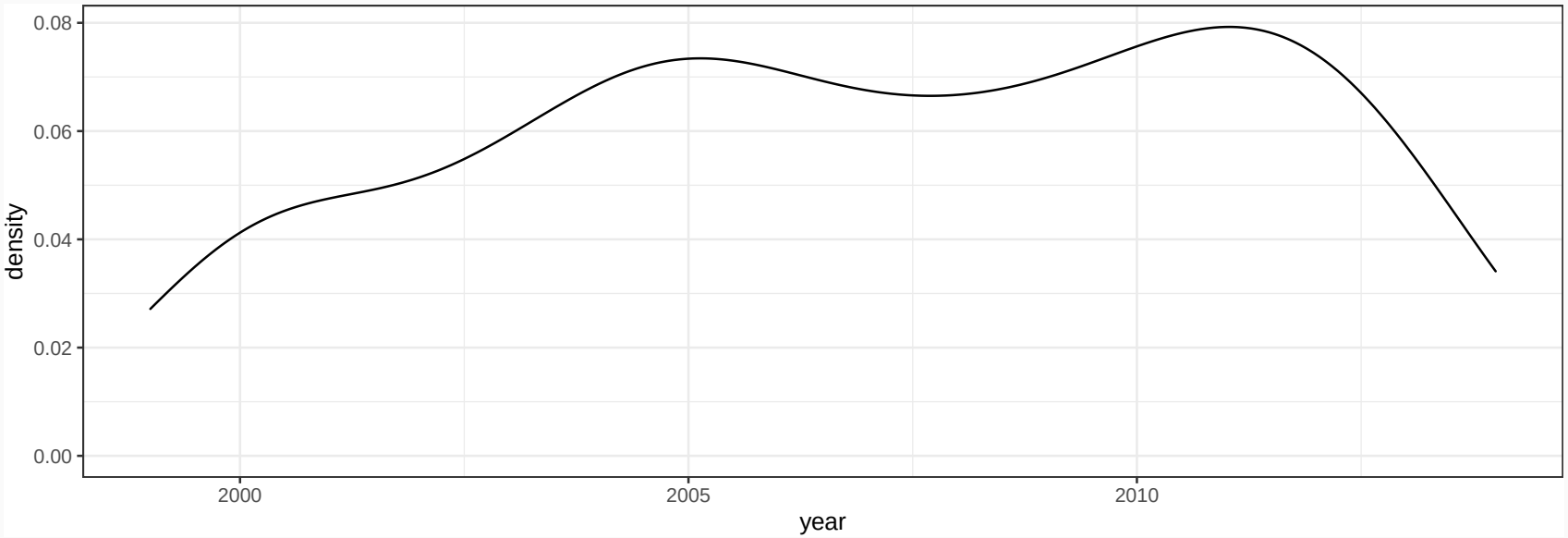
```
hist(ipv$year)
```



Density plots

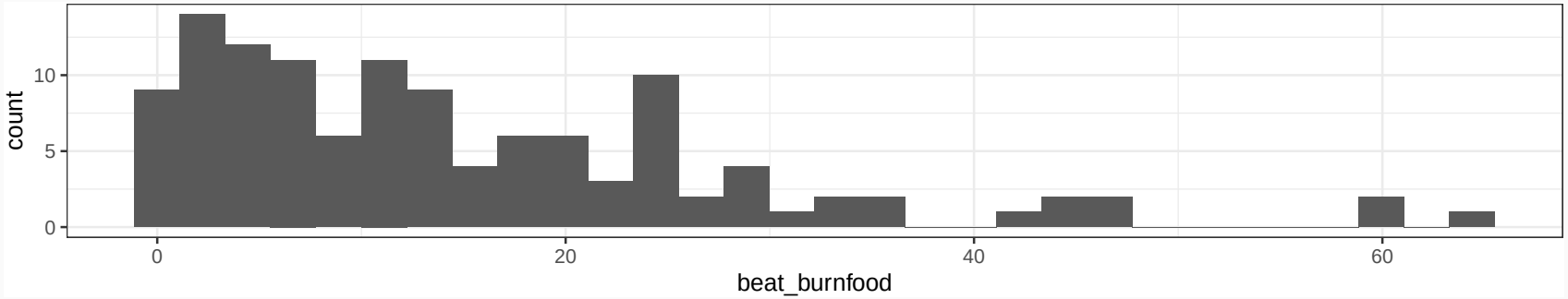
Densities are smoothed continuous histograms (with binwidth=0)

```
ggplot(ipv,
       aes(x = year)) +
  geom_density()
```



Another set of histograms/densities

```
ggplot(ipv, aes(x = beat_burnfood)) +
  geom_histogram()
```

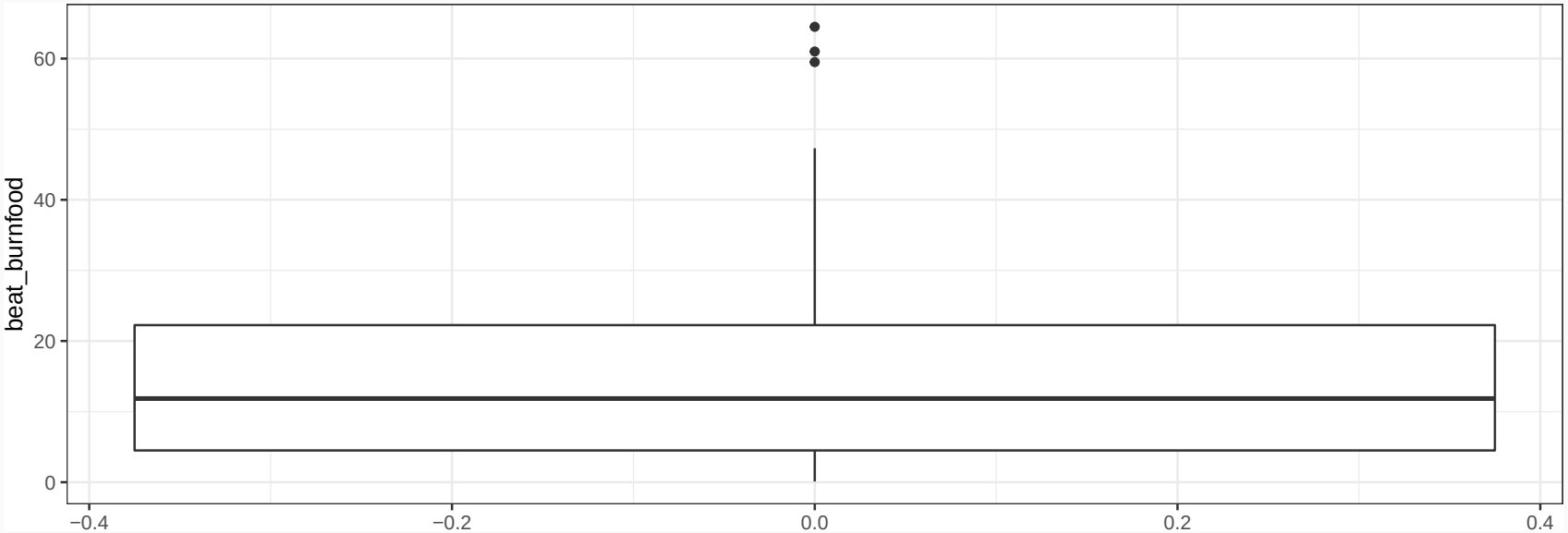



```
ggplot(ipv, aes(x = beat_burnfood)) +
  geom_density()
```

Boxplots

Show the distribution of a continuous variable with the median, quartiles, and outliers

```
ggplot(ipv, aes(y = beat_burnfood)) +
  geom_boxplot()
```

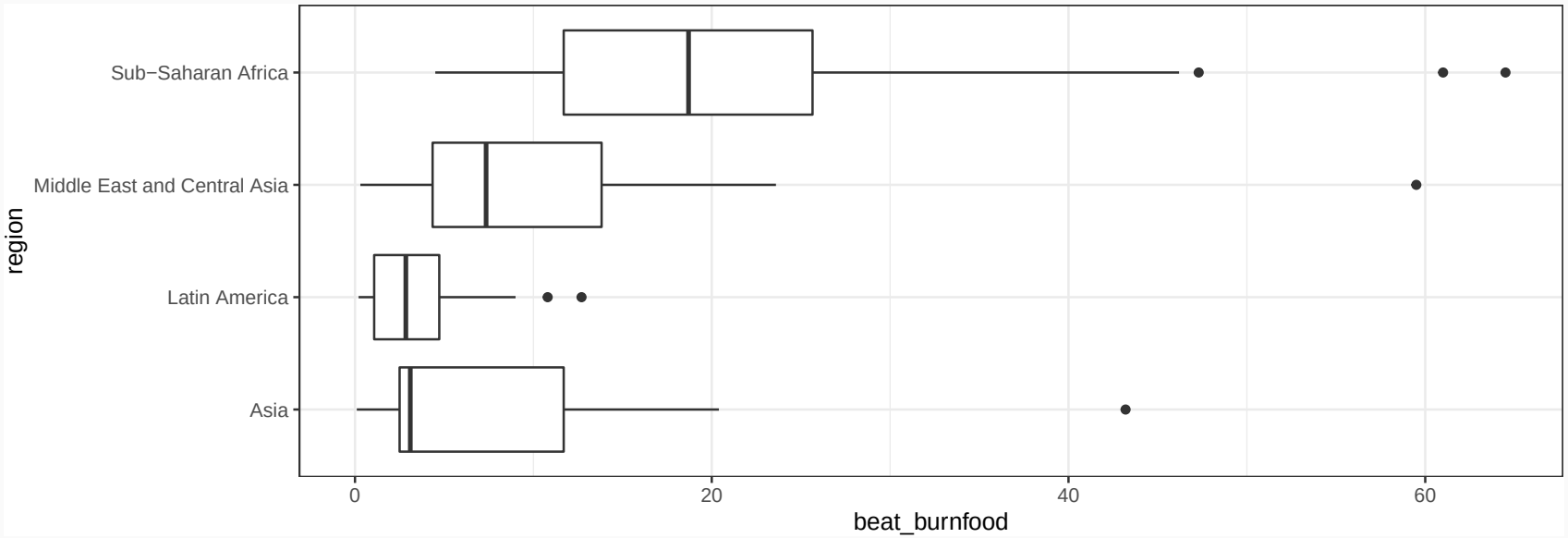


Bivariate (two variable) data visuals

One continuous, one categorical variable

Boxplots of a continuous by a categorical

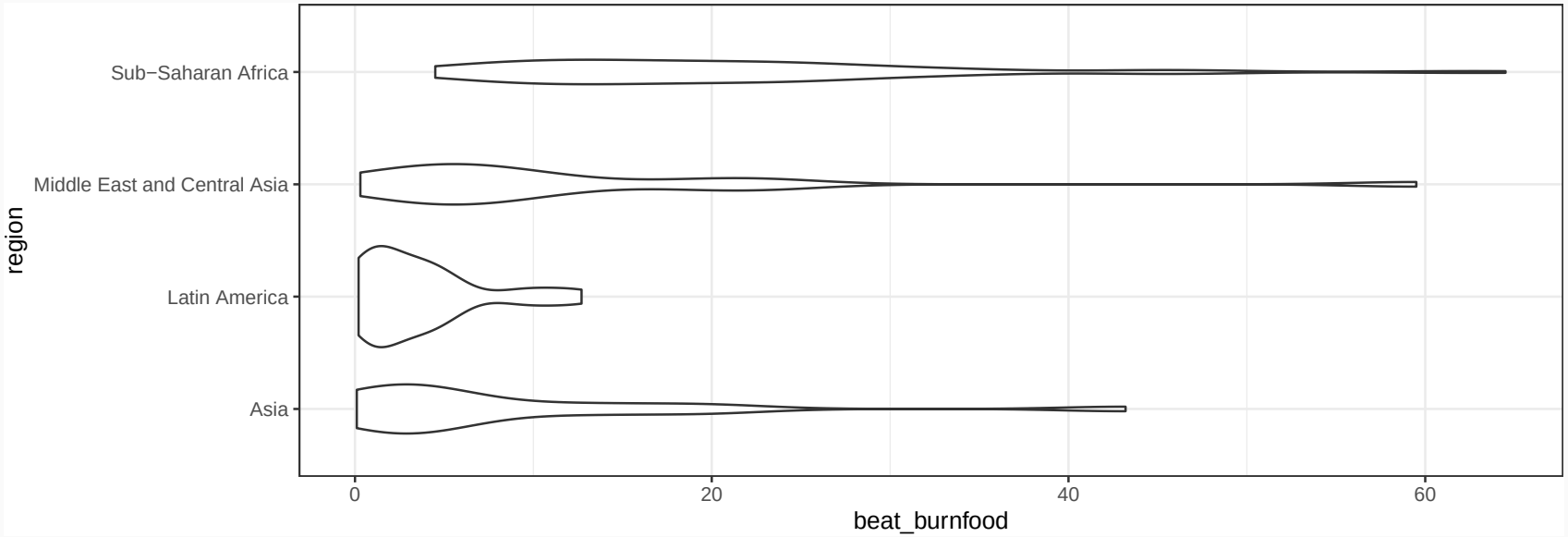
```
ggplot(ipv, aes(y = beat_burnfood, x = region)) +
  geom_boxplot() +
  coord_flip()
```



Violin plots

Like a boxplot + densityplot

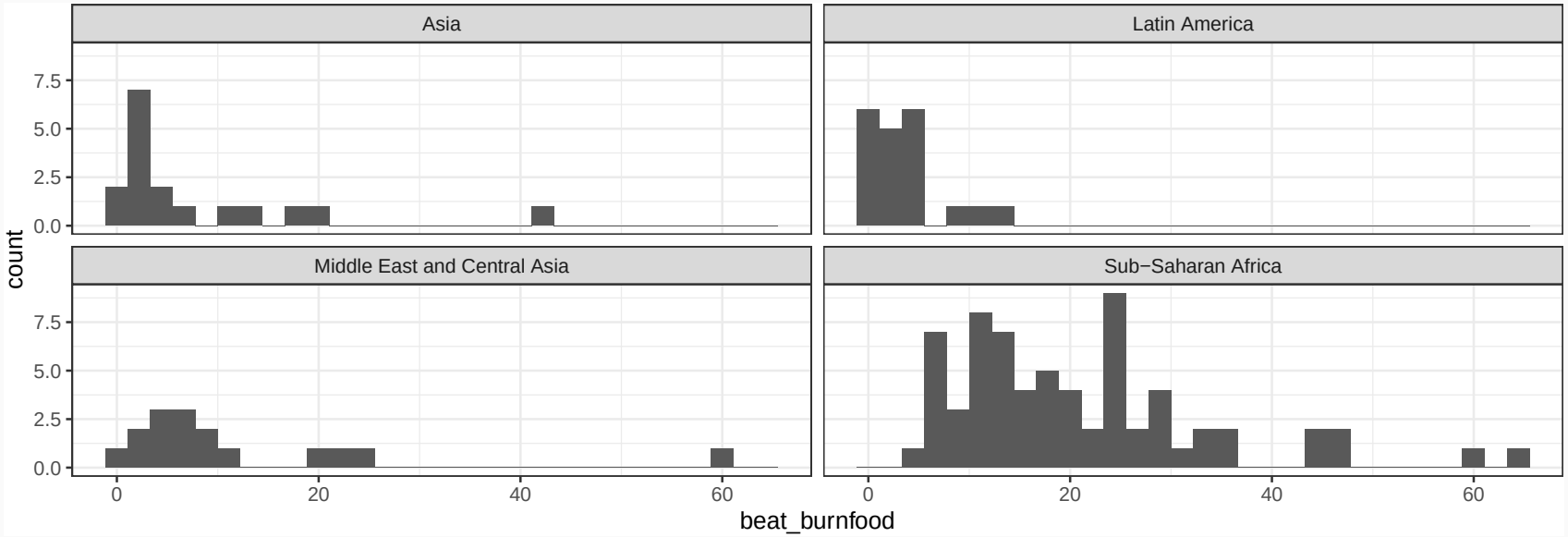
```
ggplot(ipv, aes(y = beat_burnfood, x = region)) +
  geom_violin() +
  coord_flip()
```



Faceting

Create separate plots (facet) by some categorical variable

```
ggplot(ipv, aes(x = beat_burnfood)) +
  geom_histogram() +
  facet_wrap(~region)
```

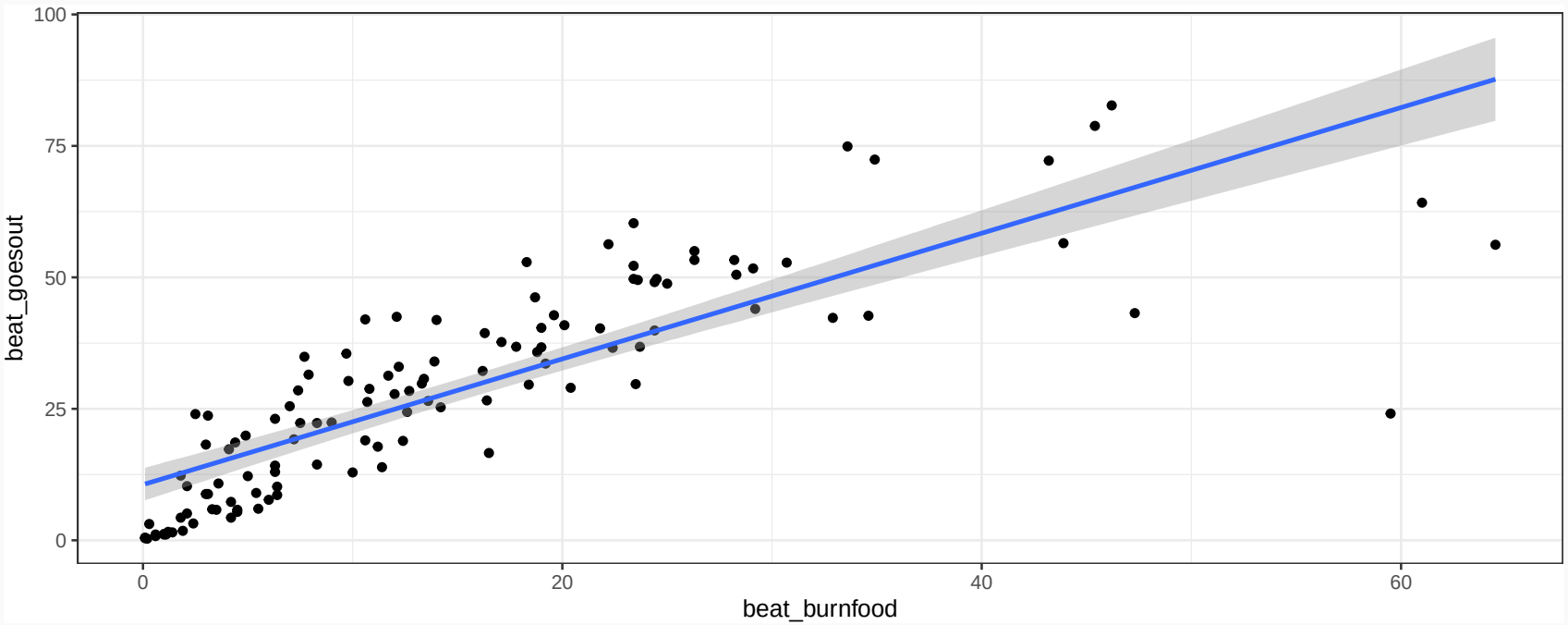


Two continuous variables

Scatterplots

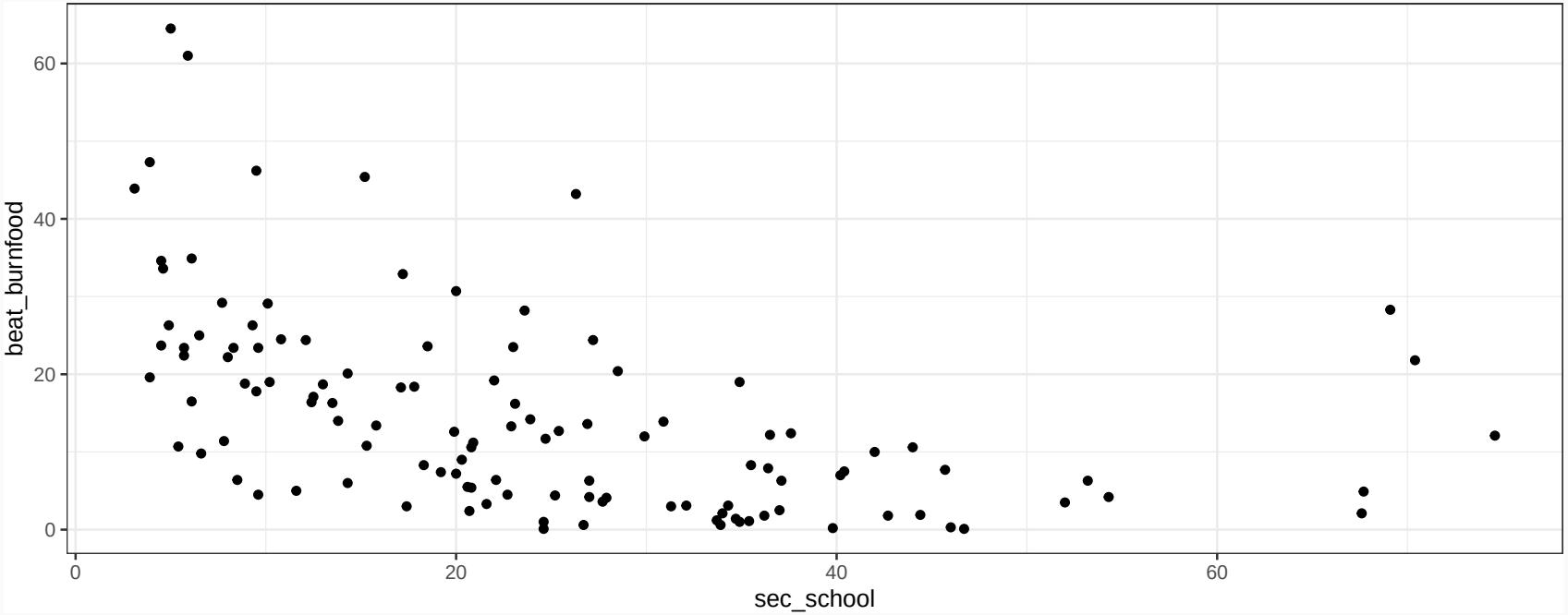
Plot points on an x,y plane based on two continuous variables

```
ggplot(ipv, aes(x = beat_burnfood, y = beat_goesout)) +
  geom_point() +
  geom_smooth(method = "lm")
```



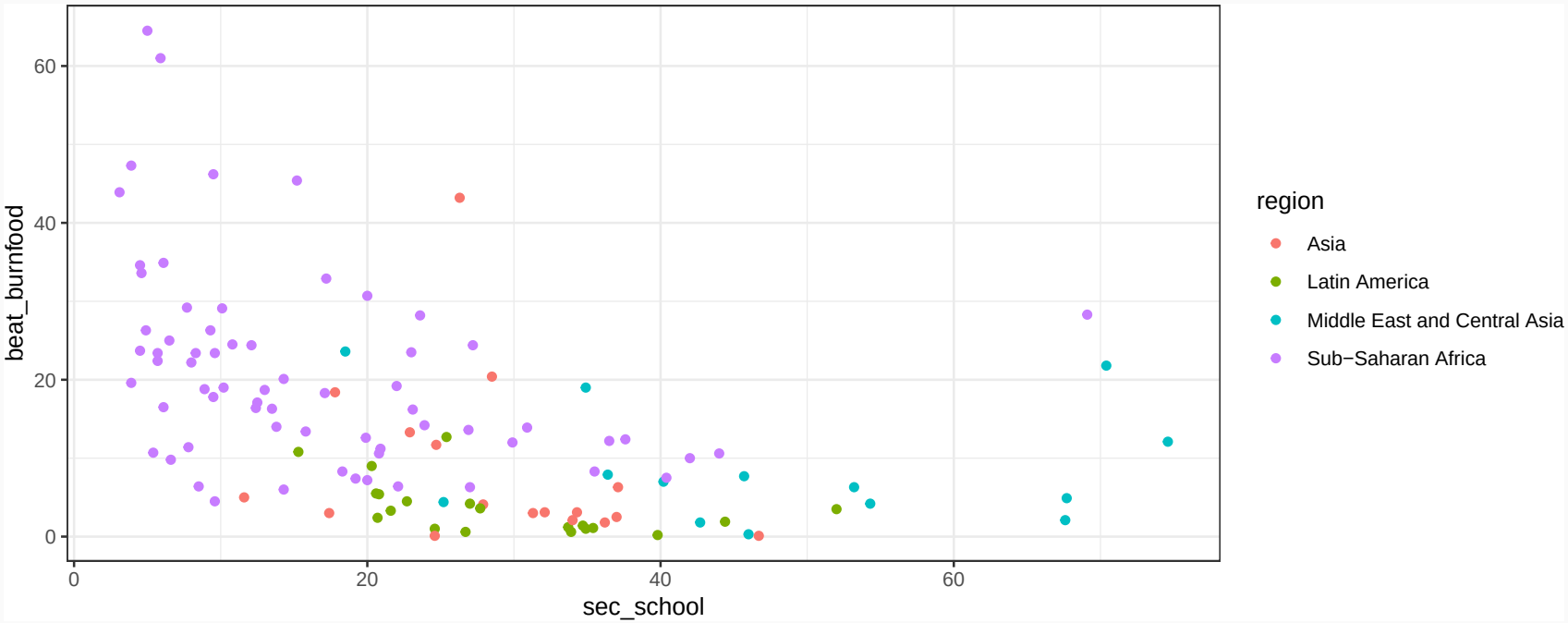
Scatterplots, continued

```
ggplot(ipv, aes(y = beat_burnfood, x = sec_school)) +
  geom_point()
```



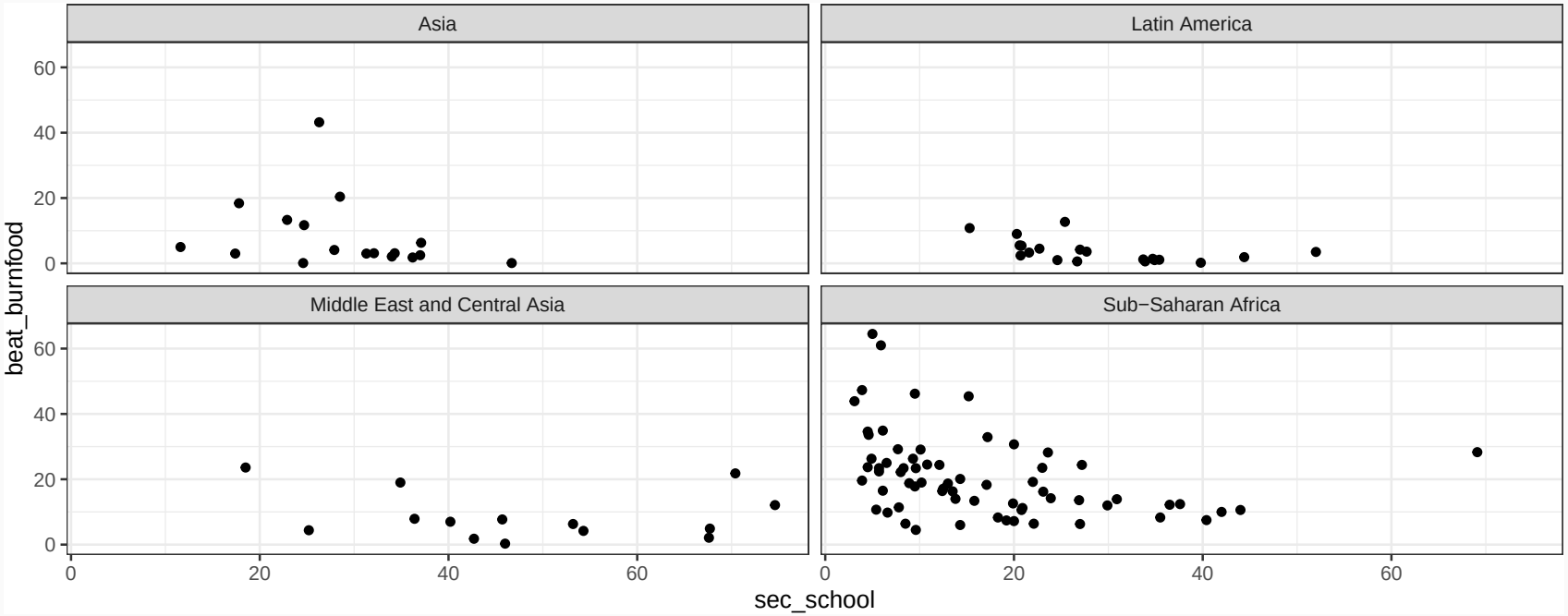
Adding a third variable to the plot

```
ggplot(ipv, aes(y = beat_burnfood, x = sec_school,
                color = region)) +
  geom_point()
```



Another way to present three variables

```
ggplot(ipv, aes(y = beat_burnfood, x = sec_school)) +
  geom_point() +
  facet_wrap(~region)
```



HW 2 Solutions review

No homework this week

- You've been doing great - take a break!
- Lab today, practice with univariate visuals in ggplot()