607 Fall 2021 Project 2 Income vs Religion

Mark Schmalfeld

10/1/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(tidyverse)
```

```
## -- 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(curl)
```

```
## Using libcurl 7.64.1 with LibreSSL/2.8.3

##
## Attaching package: 'curl'

## The following object is masked from 'package:readr':
##
##     parse_date
```

Various religions are evaluated for income across different sample sized and income ranges. This provides an opportunity to evaluate if there are income differences in each religion or not. Also to see how large a difference.

Based on the percentages of people in differnt income ranges - Does the income distribution vary by religion?

Down load data file from github with column type adjustment, name update and then evaluate data formating.

```
urlfile<-"https://raw.githubusercontent.com/schmalmr/607_Fall_2021_Project_2_Income_Religion/main/Religi
income <- read_csv(url(urlfile), col_names= c("religion","A_less_than_USD30K","B_USD30K_less_than_50K",
   religion = col_character(),
   A_less_than_USD30K = col_number(),
   B_USD30K_less_than_50K = col_number(),
   C_USD50K_less_than_100K = col_number(),
   D_USD100K_or_more = col_number(),
   sample_size =col_number()
   ))

income<-as_tibble(income)
income<-slice(income, 2:13)
glimpse(income)
```

```
## Rows: 12
## Columns: 6
## $ religion                <chr> "Buddhist", "Catholic", "Evangelical Protestan~
## $ A_less_than_USD30K      <dbl> 36, 36, 35, 17, 53, 48, 16, 29, 27, 34, 18, 33
## $ B_USD30K_less_than_50K  <dbl> 18, 19, 22, 13, 22, 25, 15, 20, 20, 17, 17, 20
## $ C_USD50K_less_than_100K <dbl> 32, 26, 28, 34, 17, 22, 24, 28, 33, 29, 36, 26
## $ D_USD100K_or_more       <dbl> 13, 19, 14, 36, 8, 4, 44, 23, 20, 20, 29, 21
## $ sample_size             <dbl> 233, 6137, 7462, 172, 1704, 208, 708, 5208, 59~
```

Transform dataset to a tidy dataset using gather

```
income_transform<-income
income_transform<-income_transform %>%
  gather("A_less_than_USD30K","B_USD30K_less_than_50K","C_USD50K_less_than_100K", "D_USD100K_or_more", 
income_transform<-arrange(income_transform,desc("sample_size"))
income_low<-filter(income_transform,income_range=="A_less_than_USD30K")
income_low<-arrange(income_low,desc(percent_sample_size))
income_high<-filter(income_transform,income_range=="D_USD100K_or_more")
income_high<-arrange(income_high,desc(percent_sample_size))
income_black<-filter(income_transform,religion=="Historically Black Protestant")
income_jewish<-filter(income_transform,religion=="Jewish")
income_catholic<-filter(income_transform,religion =="Catholic")
(summary(income_jewish)) #jewish distribution
```

```
##    religion          sample_size  income_range       percent_sample_size
##  Length:4           Min.   :708   Length:4           Min.   :15.00
##  Class :character   1st Qu.:708   Class :character   1st Qu.:15.75
##  Mode  :character   Median :708   Mode  :character   Median :20.00
##                     Mean   :708                      Mean   :24.75
##                     3rd Qu.:708                      3rd Qu.:29.00
##                     Max.   :708                      Max.   :44.00
```

```
(summary(income_catholic)) #catholic distribution
```

```
##    religion          sample_size   income_range       percent_sample_size
##  Length:4           Min.   :6137   Length:4           Min.   :19.0
##  Class :character   1st Qu.:6137   Class :character   1st Qu.:19.0
```

```
##  Mode  :character   Median :6137   Mode  :character   Median :22.5
##                     Mean   :6137                     Mean   :25.0
##                     3rd Qu.:6137                     3rd Qu.:28.5
##                     Max.   :6137                     Max.   :36.0
```

Chart the sample size vs religion. Chart the income distribution vs the religion to look at the differences in income vs religion.

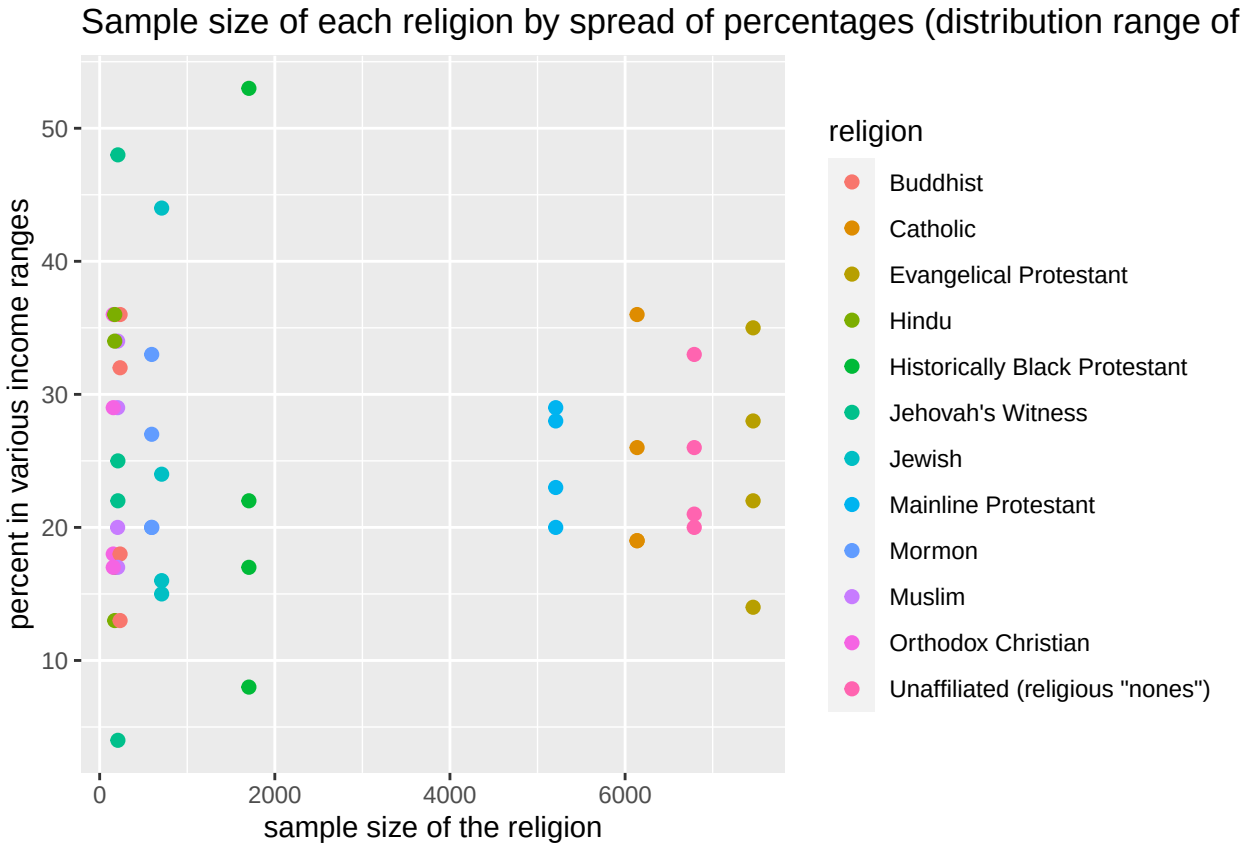
The religion sample sizes are significantly different between the various religions. The Jewish, Hindu, Muslim, Jewhovah's witness, Morman and Orthodox Christian all have relatively low sample size compared to most others which are a factor of ~10X higher. This may represent a specific area or region, it may accurate represent a region but this was not included in the dataset. The results or conclusions are only useful in context of the specific sample population we are evaluation which is not includes otherwise we can not generalize the results or conclusions beyond dataset.

The results from this sample indicate following types of conclusions for the sample: Jewsish (44%) and Budhists have the highest percentage with incomes >$100K and an unweighted average suggest 20% or so overall have an earnings of $100K or more.

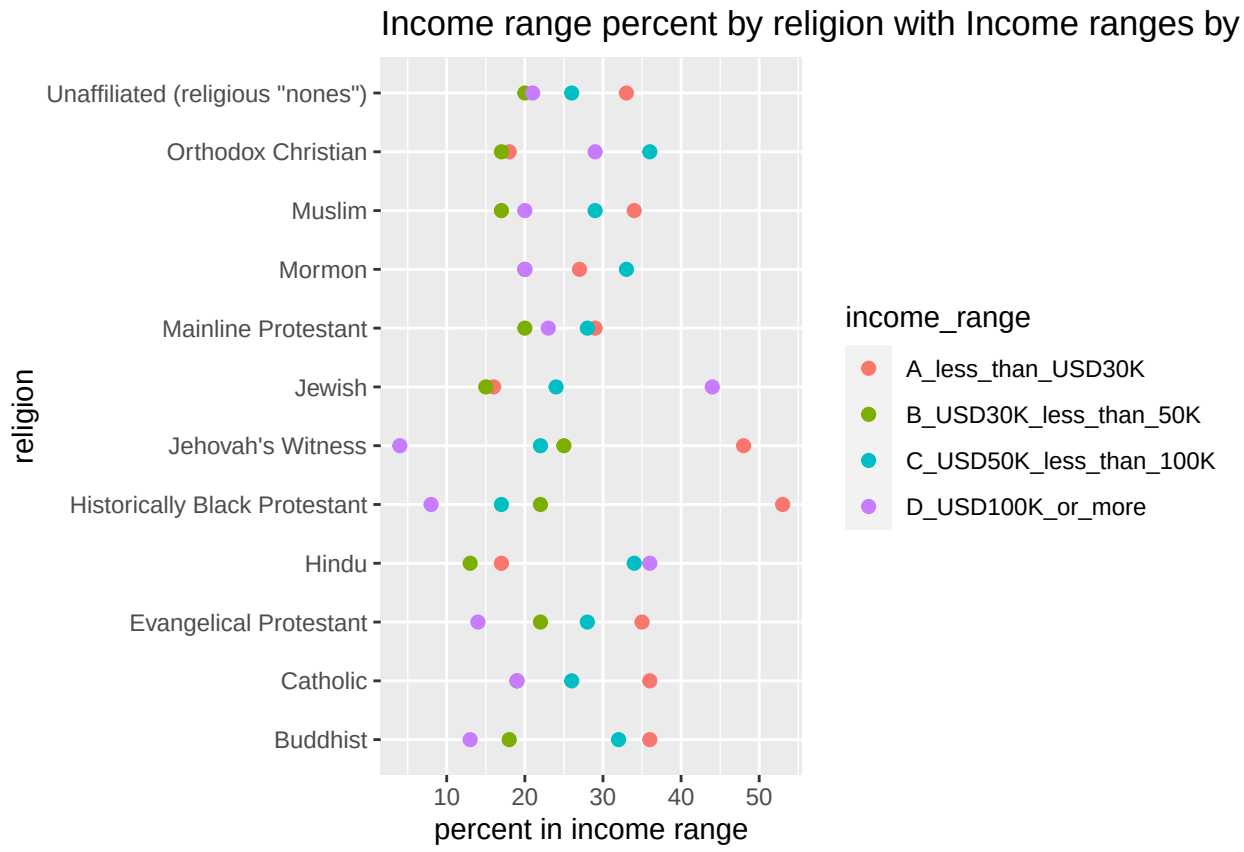
Historically Black Protestants and Jehovis witness have the highest percents with incomes less than $30K (overall in 50% range) while the next larger group is a jump down in the 35% range for the Buddhists and Muslims. It appears to be significant the poverty level differences based on religions sample - but the background reference for sample location / method could impact this a lot in the analysis and is missing.

The charts below show some of the different ways to slice the datasets for evaluation. The overall indication from the dataset sample is that religious affiliation has some relationship to income level in this data. It does not necessarily mean it is the cause of the lower income though there may a relationship to some of the religious teachings and expectations (education, learning, location of church in areas of poverty, or expectations to give heavily to the church) to the level of income or amount of giving to the church which can also impact income.

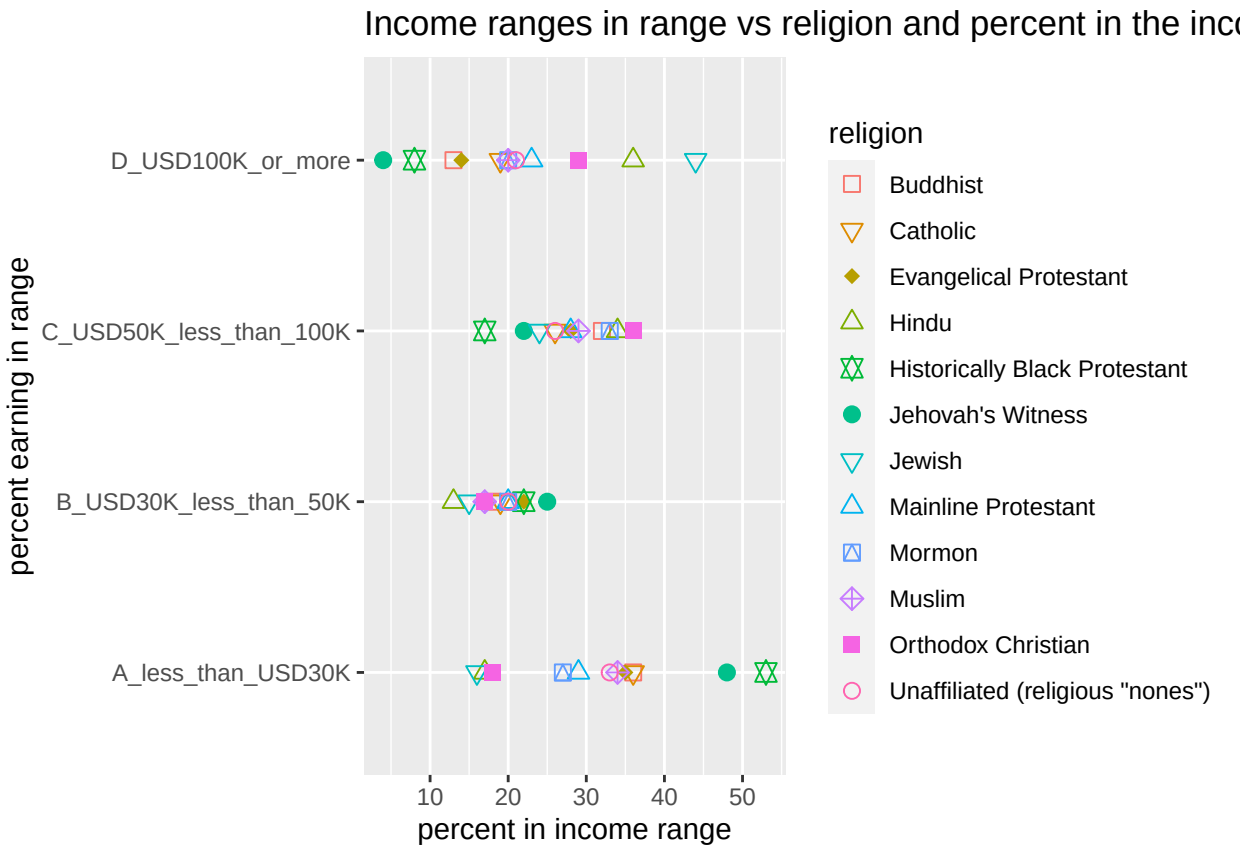
```
ggplot(income_transform,aes(x=sample_size,y=percent_sample_size,color=religion))+geom_point(size=2.0)+la
```



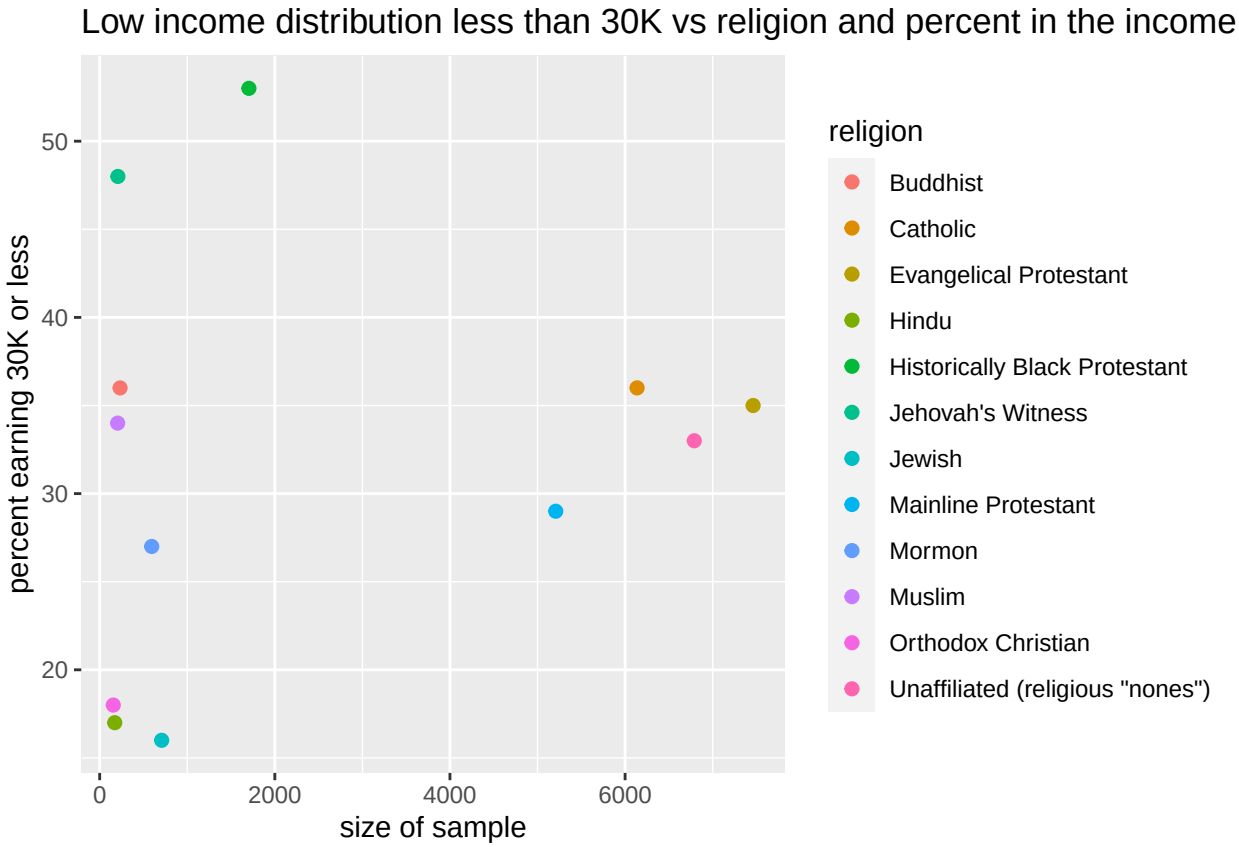

```
ggplot(income_transform,aes(x=percent_sample_size,y=religion,color=income_range))+geom_point(size=2.0)+
```



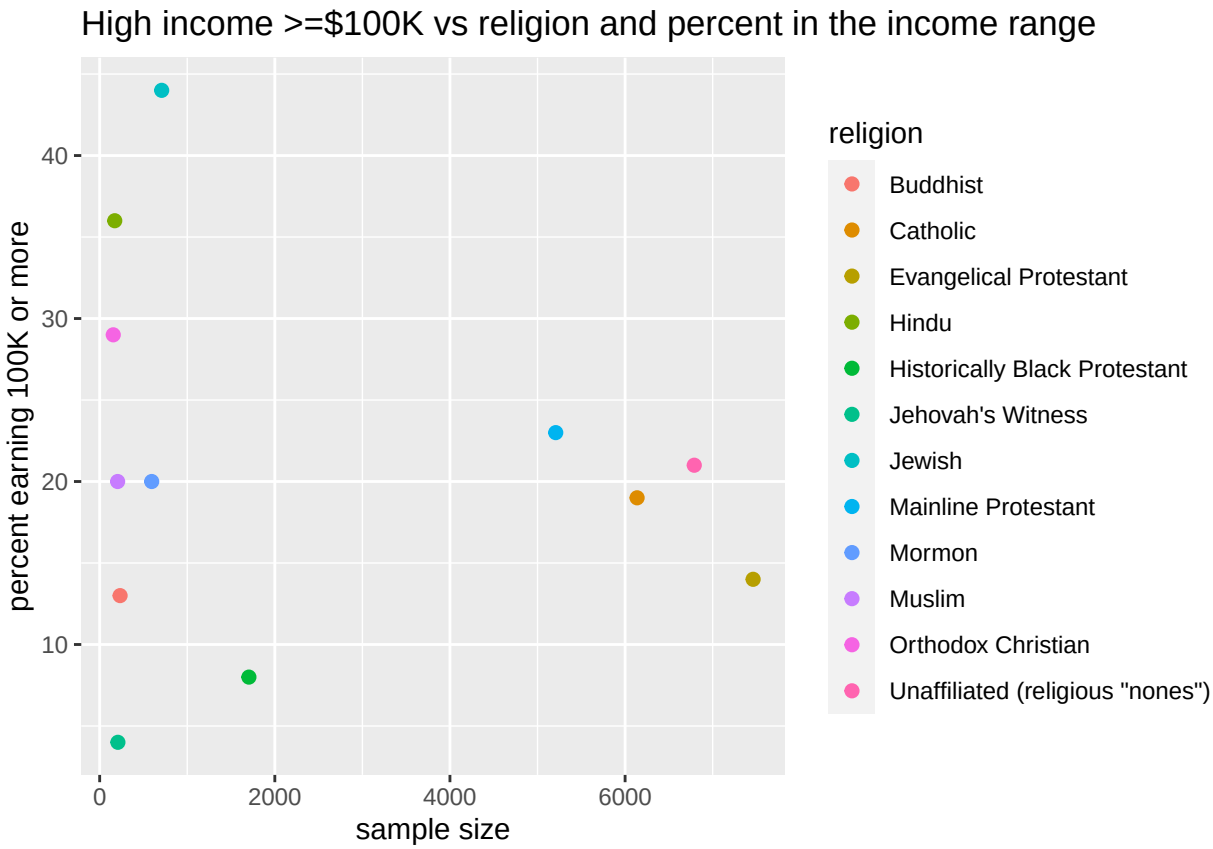

```
ggplot(income_transform,aes(x=percent_sample_size,y=income_range,shape=religion,color=religion))+geom_p
```



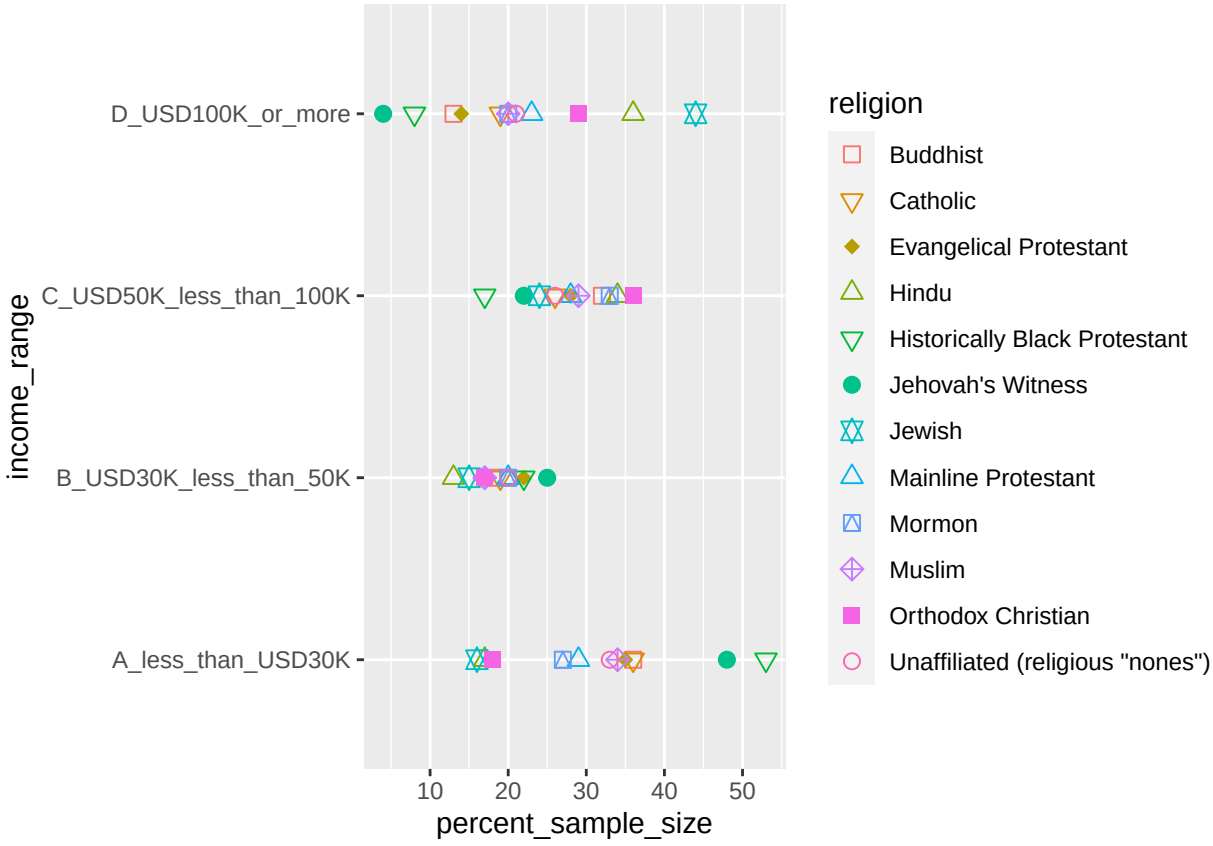

```
ggplot(income_low,aes(x=sample_size,y=percent_sample_size,color=religion))+geom_point(size=2.0)+labs(ti
```



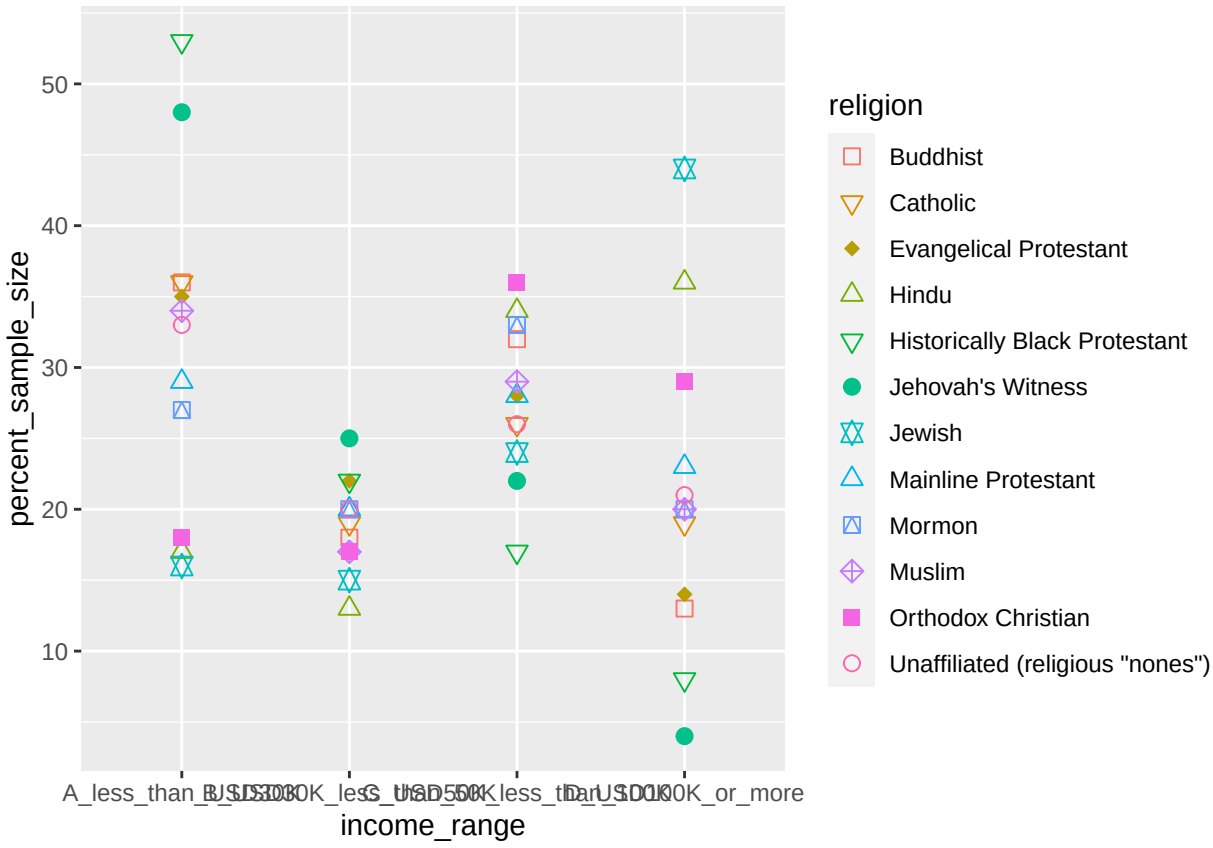

```
ggplot(income_high,aes(x=sample_size,y=percent_sample_size,color=religion))+geom_point(size=2.0)+labs(t
```




```
ggplot(income_transform,aes(y=income_range,x=percent_sample_size,shape=religion,color=religion))+geom_p
```




```
ggplot(income_transform,aes(x=income_range,y=percent_sample_size,shape=religion,color=religion))+geom_p
```


50
40
30
20
10
percent_sample_size
A_less_than_$10K
income_range
religion
Buddhist
Catholic
Evangelical Protestant
Hindu
Historically Black Protestant
Jehovah's Witness
Jewish
Mainline Protestant
Mormon
Muslim
Orthodox Christian
Unaffiliated (religious "nones")