Neighborhood Social Processes Descriptives

Gabriela Suarez

11 March, 2023

Install & Load Packages

```
library(readr)
library(foreign)
library(plyr)
library(tidyverse)
library(psych)
library(labelled)
library(ggplot2)
library(plotly)
```

Load Data

```
mtwins = read.spss(file = 'mtwins_ecv_faces_amy.sav', header = F, to.data.frame=TRUE, use.value.labels = F)
nis = read.spss(file = 'Suarez NIS data 05172021.sav', header = T, to.data.frame=TRUE, use.value.labels = F)
```

Demographics

Read Value Labels

```
as.data.frame(attr(mtwins$twin_sex, 'value.labels'))
```

	attr(mtwins$twin_sex, "value.labels") <dbl>
Female	1
Male	0

2 rows

```
as.data.frame(attr(mtwins$pc_yrace, 'value.labels'))
```

	attr(mtwins$pc_yrace, "value.labels") <dbl>
88 Missing	88
77 Refused	77
6 Other	6
5 Biracial	5
4 Asian/Pacific Islander	4
3 Native American/Native Alaskan	3
2 Black/African American	2
1 Hispanic/Latino	1

	attr(mtwins$pc_yrace, "value.labels") <dbl>
0 White/Caucasian	0
-99 N/A	-99

1-10 of 10 rows

```
as.data.frame(attr(mtwins$pc_education, 'value.labels'))
```

8 Graduate professional training, graduate degree
7 Standard college or university graduation (4 years)
6 Junior college/associates degree (2 years)
5 Partial college (at least one year) or specialized training
4 High school graduate/GED
3 Partial high school (at least one year)
2 Junior high/middle school completed
1 7th grade or pcless
0 No formal schooling

9 rows | 1-1 of 2 columns

```
as.data.frame(attr(mtwins$pc_annincome, 'value.labels'))
```

	attr(mtwins$pc_annincome, "value.labels") <dbl>
12 $90000 or more	12
11 $80000-$89000	11
10 $70000-$79999	10
9 $60000-$69999	9
8 $50000-$59999	8
7 $40000-$49999	7
6 $30000-$39999	6
5 $25000-29999	5
4 $20000-$24999	4
3 $15000-$19000	3

1-10 of 13 rows Previous **1** 2 Next

Define Missing Values

```
mtwins[mtwins == 77] = NA
mtwins[mtwins == 88] = NA
mtwins[mtwins == -99] = NA
```

Descriptives

```
table(mtwins$twin_sex)
```

```
##
##   0   1
## 386 322
```

```
demos = select(mtwins, twin_age_yr, pc_yrace, pc_education, pc_annincome)
describe(demos)
```

	vars <int>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>	trimmed <dbl>	mad <dbl>	min <dbl>	max <dbl>
twin_age_yr	1	708	14.1412429	2.238242	14	14.3098592	1.4826	7	19
pc_yrace	2	708	0.7853107	1.604170	0	0.3485915	0.0000	0	6
pc_education	3	706	6.2917847	1.301332	7	6.3745583	1.4826	1	8
pc_annincome	4	694	9.3227666	3.012169	10	9.7931655	2.9652	0	12

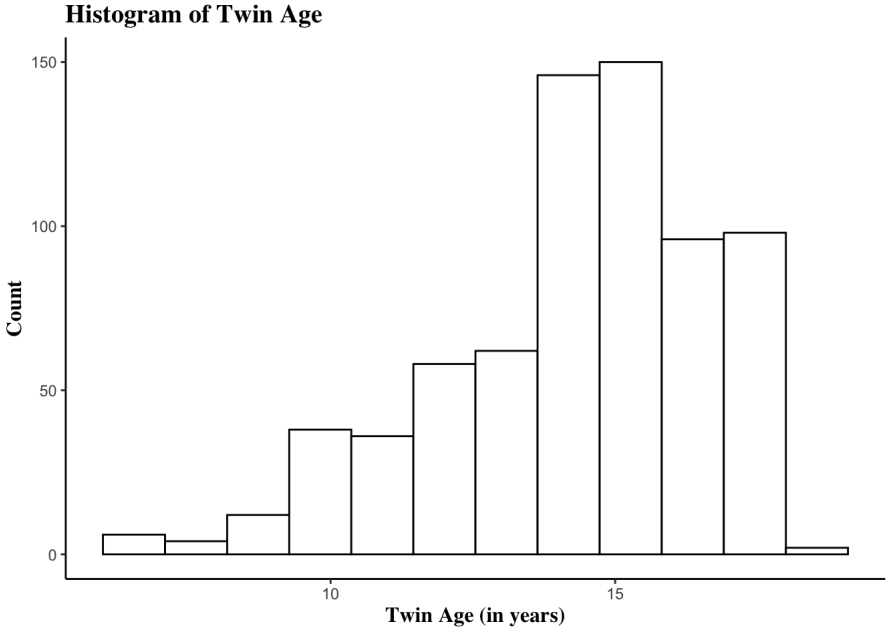
4 rows | 1-10 of 14 columns

```
apply((demos), 2, table)
```

```
## $twin_age_yr
##
##   7   8   9  10  11  12  13  14  15  16  17  18  19
##   6   4  12  38  36  58  62 146 150  96  80  18   2
##
## $pc_yrace
##
##   0   1   2   3   4   5   6
## 538  10  84   4   4  58  10
##
## $pc_education
##
##   1   3   4   5   6   7   8
##   2   8  52 156 122 232 134
##
## $pc_annincome
##
##   0   1   2   3   4   5   6   7   8   9  10  11  12
##   4   6  18  16  18  14  52  66  38  70  46  82 264
```

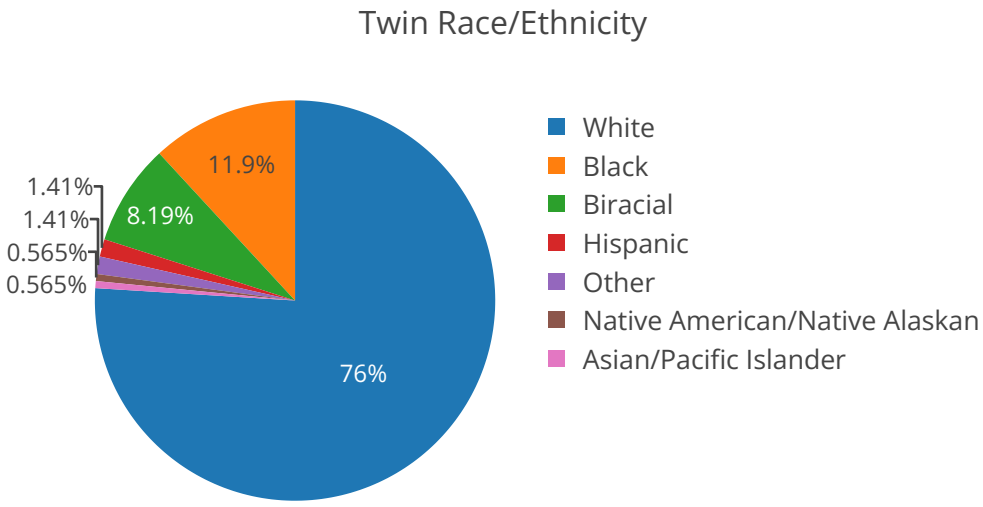
Twin Age: Histogram

```
mtwins %>%
  ggplot(aes(x = twin_age_yr)) +
  geom_histogram(bins = 12, fill = 'white', col = 'black') +
  labs(title = "Histogram of Twin Age",
       x = "Twin Age (in years)", y = "Count") +
  theme_classic() + theme(axis.title.x = element_text(size = 12, family = "Times"),
                          axis.title.y = element_text(size = 12, family = "Times"),
                          title = element_text(size = 12, face = 'bold', family = 'Times'))
```



Twin Race: Pie Chart

```
race = data.frame(
  group = c("White","Hispanic","Black","Native American/Native Alaskan","Asian/Pacific Islander","Biracial","Othe
r"),
  value = c(538, 10, 84, 4, 4, 58, 10)
)

# Create plotly interactive pie chart
fig = plot_ly(race,
              labels = ~group,
              values = ~value,
              type = 'pie',
              textfont=list(size=18))

fig = fig %>% layout(title = list(text='Twin Race/Ethnicity', font=list(size=24)),
         xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE, font=list(size=18)),
         yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE, font=list(size=18)),
         showlegend=TRUE, legend=list(font=list(size=20)))

m = list(l = 50, r = 50, b = 100, t = 100, pad = 4)

fig = fig %>% layout(autosize = F, width = 750, height = 500, margin=m)
fig
```



Neighborhood Disadvantage

Neighborhood disadvantage is defined using geocoding of family addresses to assess the proportion of neighborhood residents living below the poverty line (10.5% was the average at the time of data collection) between 2011 and 2015 in each family's census tract.

```
describe(mtwins$pov1115)
```

	vars	n	mean	sd	median	trimmed	mad	min	max
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
X1	1	708	0.1962339	0.1600671	0.1515	0.1733725	0.1245384	0	0.77

1 row | 1-10 of 14 columns

```
# How many families are living in neighborhoods where more than 10.5% of families in the neighborhood are living
below the poverty line?
mtwins$belowPovLine <- ifelse(mtwins$pov1115 > .105, '>10.5% living below pov line', '<10.5% living below pov lin
e')
table(mtwins$belowPovLine)
```

```
##
## <10.5% living below pov line >10.5% living below pov line
##                          254                          454
```

```
prop.table(table(mtwins$belowPovLine))
```

```
##
## <10.5% living below pov line >10.5% living below pov line
##                    0.3587571                    0.6412429
```

Michigan Census tract data

Load Data

```
michigan = read_csv("/Users/gabrielasuarez/Dropbox (University of Michigan)/MiND_Lab/Research_Projects/nhood_soci
al_processes/Analysis/ACS_allCTs_2008-2012_families.csv")
```

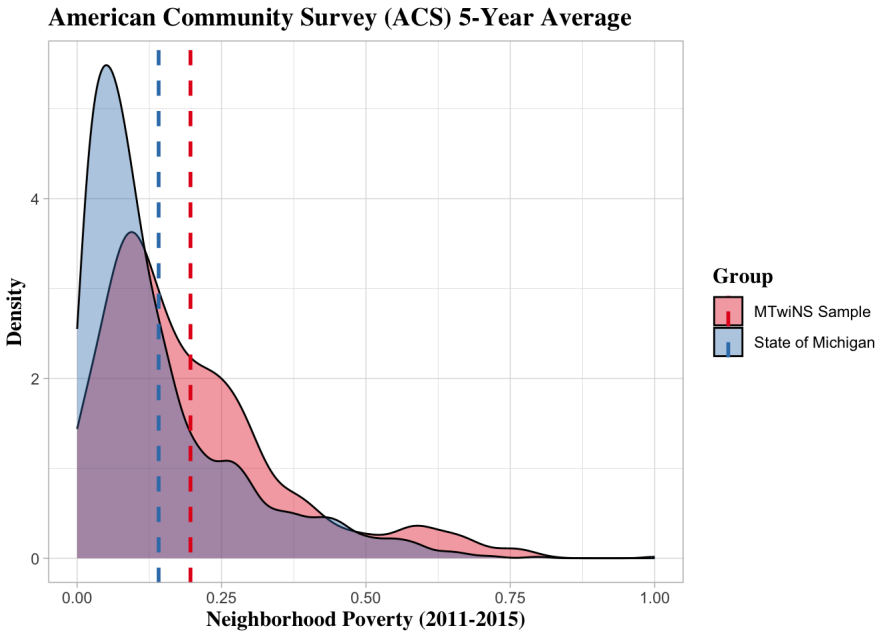
Prep Data

```
# Rename the pov_2015 variable to pov1115
michigan$pov1115 = michigan$DP03_0119PE
michigan$pov1115 = michigan$pov1115/100
```

```
# Create two new dataframes only with the data that I need
michigan_dat = dplyr::select(michigan, pov1115)
mtwins_dat = dplyr::select(mtwins, pov1115)

# Make a new column in each dataframe that will be a variable to identify where they came from later
mtwins_dat$Group = 'MTwiNS Sample'
michigan_dat$Group = 'State of Michigan'

# Combine into new dataframe "nhood_pov"
nhoodPov = rbind(mtwins_dat, michigan_dat)

# Calculate the mean in each group
mu = ddply(nhoodPov, "Group", summarise, grp.mean=mean(pov1115, na.rm=T))
```

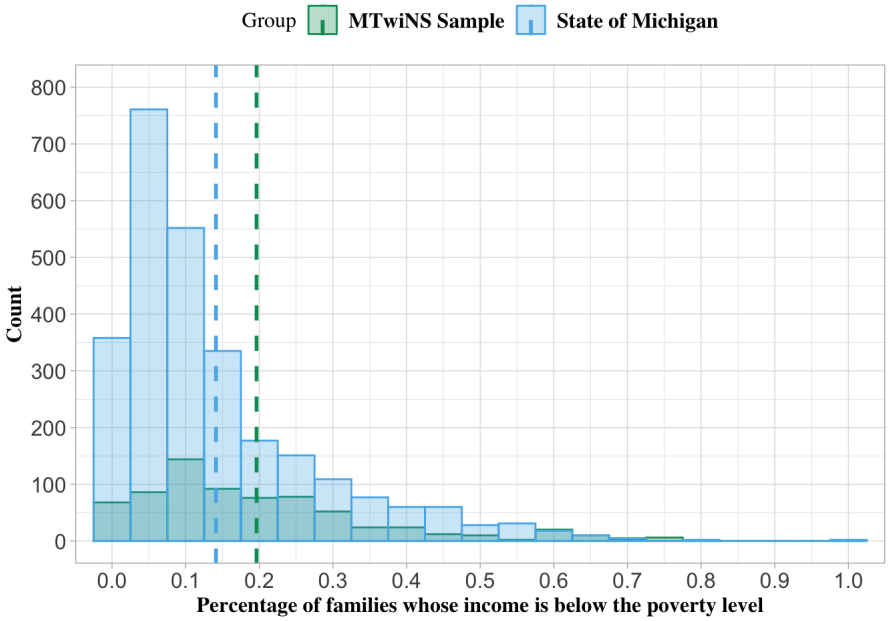
Overlaid Density Plots

```
ggplot(nhoodPov, aes(pov1115, fill = Group)) +
  geom_density(alpha = 0.4, position = 'identity') +
  geom_vline(data=mu, aes(xintercept=grp.mean, color = Group), linetype="dashed", size= 1) +
  scale_color_brewer(palette="Set1") +
  scale_fill_brewer(palette="Set1") +
  labs(title = "American Community Survey (ACS) 5-Year Average",
       x = "Neighborhood Poverty (2011-2015)", y = "Density")+
  theme_light() + theme(axis.title.x = element_text(size = 12, family = 'Times'),
                        axis.title.y = element_text(size = 12, family = 'Times'),
                        title = element_text(size = 12, face = 'bold', family = 'Times'))
```



Overlaid Histograms

```
palette1 = c("#009966","#56B4E9")

plot = ggplot(nhoodPov, aes(x = pov1115, fill = Group, color = Group)) +
  geom_histogram(binwidth = 0.05, alpha = 0.3, position = 'identity') +
  geom_vline(data=mu, aes(xintercept=grp.mean, color = Group), linetype="dashed", size=1) +
  theme(legend.position="bottom") +
  scale_color_manual(values=palette1) +
  scale_fill_manual(values=palette1) +
  scale_x_continuous(name="Percentage of families whose income is below the poverty level", breaks = seq(0, 1, 0.
1)) +
  scale_y_continuous(name="Count", breaks = seq(0, 800, 100)) +
  coord_cartesian(xlim=c(0, 1), ylim=c(0, 800)) +
  #labs(title = "American Community Survey (ACS) 5-Year Average")+
  theme_light() + theme(axis.title.x = element_text(size = 12, face = "bold", family="Times"),
                        axis.title.y = element_text(size = 12, face = "bold", family="Times"),
                        legend.position = "top",
                        legend.title = element_text(size = 12, face = "plain", family="Times"),
                        axis.text.x = element_text(size = 12),
                        axis.text.y = element_text(size = 12),
                        legend.text = element_text(size = 12, face = "bold", family="Times"))
plot
```



```
#ggsave("neighborhood_dist.png", plot=plot, width = 30, height = 20, unit = "cm", dpi=600)
```

Overlaid Histogram & Density Plot

```
ggplot(nhoodPov, aes(x = pov1115, fill = Group, color = Group))+
  geom_histogram(binwidth = 0.05, alpha = 0.6, aes(y = ..density..), position = 'identity')+
  geom_vline(data=mu, aes(xintercept=grp.mean, color = Group), linetype="dashed")+
  geom_density(alpha=0.1)+
  scale_color_brewer(palette="Set1")+
  scale_fill_brewer(palette="Set1")+
  labs(title = "American Community Survey (ACS) 5-Year Average",
       x = "Neighborhood Poverty Rates (2011-2015)", y = "Density/Count")+
  theme_light() + theme(axis.title.x = element_text(size = 12, family = 'Times'),
                        axis.title.y = element_text(size = 12, family = 'Times'),
                        title = element_text(size = 12, face = 'bold', family = 'Times'))
```