Heat (and hexagon) plots in Stata

Ben Jann

University of Bern, ben.jann@soz.unibe.ch

2019 German Stata Users Group meeting
Munich, May 24, 2019
Outline

1 Introduction

2 Syntax of heatplot and hexplot

3 Examples
   - Bivariate histogram
   - Trivariate distributions
   - Display values as marker labels
   - Correlation matrix
   - Spacial weights matrix

4 Installation
What is a heat plot?

- Generally speaking, a heat plot is a graph in which some aspect of the data is displayed as a **color gradient**.

- A simple example is a **bivariate histogram**; the color gradient is used to illustrate (relative) frequencies within bins of $X$ and $Y$. 
. quietly drawnorm y x, n(10000) corr(1 .5 1) cstorage(lower) clear
. heatplot y x, backfill colors(plasma)
What about hexagons?

- Hexagons are great because they look a bit like circles, but you can join them together without leaving gaps.
- Bees found out how awesome hexagons are long time ago.
What about hexagons?

Latter on, gully cover designers found out that hexagons look great on gully covers.
Finally, also statisticians discovered the virtues of hexagons.

“The here are many reasons for using hexagons, at least over squares. Hexagons have symmetry of nearest neighbors which is lacking in square bins. Hexagons are the maximum number of sides a polygon can have for a regular tesselation of the plane, so in terms of packing a hexagon is 13% more efficient for covering the plane than squares. This property translates into better sampling efficiency at least for elliptical shapes. Lastly hexagons are visually less biased for displaying densities than other regular tesselations. For instance with squares our eyes are drawn to the horizontal and vertical lines of the grid.”

---

Example from above using hexagons

```
. hexplot y x, backfill colors(plasma)
```

Ben Jann (University of Bern)

heatplot

Munich, 24.05.2019
1 Introduction

2 Syntax of heatplot and hexplot

3 Examples
- Bivariate histogram
- Trivariate distributions
- Display values as marker labels
- Correlation matrix
- Spacial weights matrix

4 Installation
Main commands

- Bivariate histogram
  \texttt{heatplot} \texttt{Y \ X [if] [in] [weight] [\ , options]}

- Trivariate heat plot (color gradient for \textit{Z})
  \texttt{heatplot} \texttt{Z \ Y \ X [if] [in] [weight] [\ , options]}

- Heat plot from Stata matrix
  \texttt{heatplot} \texttt{matname [\ , options]}

- Heat plot from Mata matrix
  \texttt{heatplot mata(name) [\ , options]}

- Heat plot using hexagons
  \texttt{hexplot ...}
Main options

- **Color gradient options**
  - `levels(#)`: number of color bins
  - `cuts(numlist)`: custom cutpoints for color bins
  - `colors(palette)`: color map to be used for the color bins
  - `statistic(stat)`: how $Z$ is aggregated
  - `size[(exp)] | sizeprop`: size of color fields
  - `values(options)`: display values as marker labels
  - `scatter(...)`: render color fields as scatter plot
  - `keylabels(spec)`: how legend keys are labeled

- **Binning of $Y$ and $X$**
  - `[x|y]bins(spec)`: how continuous $Y$ and $X$ are binned
  - `[x|y]bwidth(spec)`: alternative to `bins()`
  - `[x|y]discrete[#]`: treat variables as discrete and omit binning

(note: categorical $X$ and $Y$ can be specified as `i.varname`)
Main options

- **Matrix options**
  - `drop(numlist)`
  - `lower`
  - `nodiagonal`

- **Graph options**
  - `addplot(plots)`
  - `by(varlist[, options])`
  - `twoway_options`

- Some more options related to storing results...
Introduction

2 Syntax of heatplot and hexplot

3 Examples
- Bivariate histogram
- Trivariate distributions
- Display values as marker labels
- Correlation matrix
- Spacial weights matrix

4 Installation
webuse nhanes2, clear
heatplot weight height
. heatplot weight height, xbins(20) ybwidth(10 30)
Use counts, change color ramp, change binning, and labeling

```
. heatplot weight height, statistic(count) color(plasma, reverse) ///
    cut(1(5)@max) keylabels(, range(1))
```
Use hexagons instead of squares

```
+heplot weight height, statistic(count) color(plasma, reverse) ///
>   cut(1(5)@max) keylabels(, range(1))
```
Scale size of hexagons by relative frequency

```
. hexplot weight height, statistic(count) color(plasma) ///
   cut(1(5)@max) keylabels(, range(1)) size
```
Scaling also available with squares

```
. heatplot weight height, statistic(count) color(plasma) ///
   >   cut(1(5)@max) keylabels(, range(1)) size
```

![Heat plot of weight vs. height with count statistic and plasma color scale](attachment:image.png)
Adding other plots

```
. hexplot weight height, statistic(count) color(plasma) ///
>   cut(1(5)@max) keylabels(, range(1)) size ///
>   addplot(lpolyci weight height, degree(1) psty(p2) lw(*1.5) ac(%50) alc(0))
```
Introduction

Syntax of heatplot and hexplot

Examples
- Bivariate histogram
- Trivariate distributions
- Display values as marker labels
- Correlation matrix
- Spacial weights matrix

Installation
Gender distribution (proportion female) by weight and height

.webuse nhanes2, clear
.hexplot female weight height, color(PiYG) ylabel(25(25)175) cuts(0(.05)1)
Same graph taking account relative frequencies

```
> hexplot female weight height, color(PiYG) ylabel(25(25)175) cuts(0(.05)1) ///
>     sizeprop recenter p(lcolor(black) lwidth(vthin) lalign(center))
```

Ben Jann (University of Bern)
Distribution of the body mass index by gender and its relation to high blood pressure

. heatplot highbp bmi i.female, xdiscrete(0.9) yline(18.5 25) cuts(0(.05).75) ///
>    sizeprop recenter colors(inferno) plotregion(color(gs11)) ylabel(, nogrid)

Body Mass Index (BMI)

1=female, 0=male
Sea surface temperature by longitude, latitude, and date

```
.sysuse surface, clear
(NOAA Sea Surface Temperature)
. heatplot temperature longitude latitude, discrete(.5) statistic(asis) ///
  by(date, legend(off)) ylabel(30(1)38) aspectratio(1)
```

Graphs by date

Ben Jann (University of Bern)
Same plot using hexagons

```
. hexplot temperature longitude latitude, discrete(.5) statistic(asis) ///
    by(date, legend(off)) ylabel(30(1)38) aspectratio(1)
```

Graphs by date

Ben Jann (University of Bern)
Same plot using hexagons

```
. hexplot temperature longitude latitude, discrete(.5) statistic(asis) clip ///
>     by(date, legend(off)) ylabel(30(1)38) aspectratio(1)
```

Graphs by date

Ben Jann (University of Bern)
1 Introduction

2 Syntax of heatplot and hexplot

3 Examples
  - Bivariate histogram
  - Trivariate distributions
  - Display values as marker labels
  - Correlation matrix
  - Spacial weights matrix

4 Installation
Same plot using hexagons

```stata
quietly sysuse auto, clear
hexplot price weight mpg, values(format(%9.0f)) legend(off) aspectratio(1) ///
>     colors(plasma, intensity(.6)) p(lc(black) lalign(center))
```

![Hexagonal plot showing weight vs. mileage with data points labeled](image-url)
1 Introduction

2 Syntax of heatplot and hexplot

3 Examples
   - Bivariate histogram
   - Trivariate distributions
   - Display values as marker labels
   - Correlation matrix
   - Spatial weights matrix

4 Installation
First store correlations in a matrix and then plot from there

```stata
. quietly sysuse auto, clear
. quietly correlate price mpg trunk weight length turn foreign
. matrix C = r(C)
. heatplot C, values(format(%9.3f)) color(hcl, diverging intensity(.6)) ///
   > legend(off) aspectratio(1)
```

![Correlation Matrix](image)
Plot lower triangle only

. heatplot C, values(format(%9.3f)) color(hcl, diverging intensity(.6)) ///
>     legend(off) aspectratio(1) lower nodiagonal

-0.469
0.314
0.539
0.432
0.310

-0.582
-0.807
-0.796
-0.719
-0.359

0.672
0.727
0.601
-0.359

0.946
0.857
-0.593
-0.570

0.864
-0.570

0.631

mpg
trunk
weight
length
turn

price
mpg
trunk
weight
length
turn

Ben Jann (University of Bern) Munich, 24.05.2019 39
1 Introduction

2 Syntax of heatplot and hexplot

3 Examples

- Bivariate histogram
- Trivariate distributions
- Display values as marker labels
- Correlation matrix
- Spatial weights matrix

4 Installation
Copy some data


Compute spacial weights matrix (this might take a while)

. use homicide1990  
(S.Messner et al.(2000), U.S southern county homicide rates in 1990)
. spmatrix create contiguity W
. spmatrix matafromsp W id = W
. mata mata describe W

<table>
<thead>
<tr>
<th># bytes</th>
<th>type</th>
<th>name and extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>15,949,952</td>
<td>real matrix</td>
<td>W[1412,1412]</td>
</tr>
</tbody>
</table>

(matrix $W$ has about 2 million cells)
Heat plot of $W$ with default settings, ignoring cells (i.e. weights) that are equal to zero

```
. heatplot mata(W), drop(0) aspectratio(1)
```
Hexagon plot with fine-grained resolution

```
. heatplot mata(W), drop(0) aspectratio(1) hexagon bins(100)
```
Plot each cell individually using the discrete option

```
. heatplot mata(W), drop(0) aspectratio(1) discrete color(black) p(lalign(center))
```
Could also use the scatter option

. heatplot mata(W), drop(0) aspectratio(1) discrete color(black) scatter p(ms(p))
1 Introduction

2 Syntax of heatplot and hexplot

3 Examples
- Bivariate histogram
- Trivariate distributions
- Display values as marker labels
- Correlation matrix
- Spacial weights matrix

4 Installation
Installation

- To install `heatplot` (and `hexplot`) type
  
  . ssc install heatplot, replace

- `heatplot` depends on the `palettes` package, which itself depends on the ColrSpace Mata library, so you may also want to type
  
  . ssc install palettes, replace
  . ssc install colrspace, replace