

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

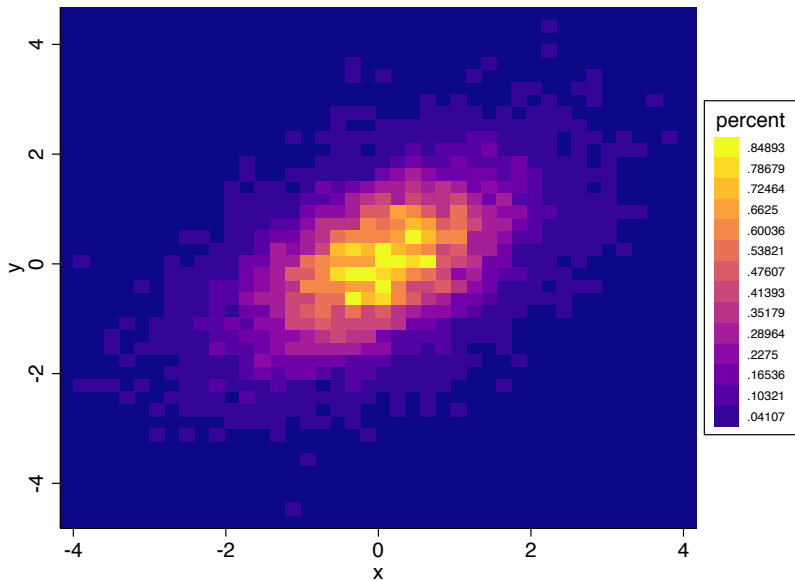
Outline

- 1 Introduction
- 2 Syntax of `heatmap` and `hexplot`
- 3 Examples
 - Bivariate histogram
 - Trivariate distributions
 - Display values as marker labels
 - Correlation matrix
 - Spatial 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  
. heatmap 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.



What about hexagons?

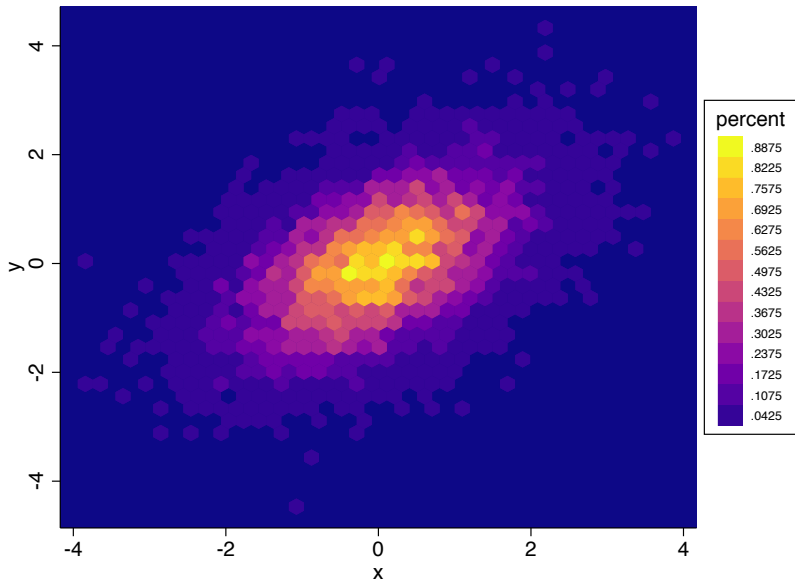
- Finally, also statisticians discovered the virtues of hexagons.

“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 tessellation 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 tessellations. For instance with squares our eyes are drawn to the horizontal and vertical lines of the grid.”¹

¹Lewin-Koh, N. (2018). Hexagon Binning: an Overview. Available from https://cran.r-project.org/web/packages/hexbin/vignettes/hexagon_binning.pdf

Example from above using hexagons

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



1 Introduction

2 Syntax of `heatmap` and `hexplot`

3 Examples

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

4 Installation

Main commands

- Bivariate histogram

```
heatplot Y X [if] [in] [weight] [, options]
```

- Trivariate heat plot (color gradient for Z)

```
heatplot Z Y X [if] [in] [weight] [, options]
```

- Heat plot from Stata matrix

```
heatplot matname [, options]
```

- Heat plot from Mata matrix

```
heatplot mata(name) [, options]
```

- Heat plot using hexagons

```
hexplot ...
```

Main options

- Color gradient options

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

- Binning of Y and X

<code>[x y]bins(spec)</code>	how continuous Y and X are binned
<code>[x y]bwidth(spec)</code>	alternative to <code>bins()</code>
<code>[x y]discrete[(#)]</code>	treat variables as discrete and omit binning
(note: categorical X and Y can be specified as <code>i.varname</code>)	
...	

Main options

- Matrix options

`drop(numlist)`

`lower`

`lower`

`nodiagonal`

drop elements equal to values in *numlist*

display lower triangle only

display upper triangle only

omit diagonal

- Graph options

`addplot(plots)`

`by(varlist [, options])`

`twoway_options`

...

add other plots to the graph

repeat plot by subgroups

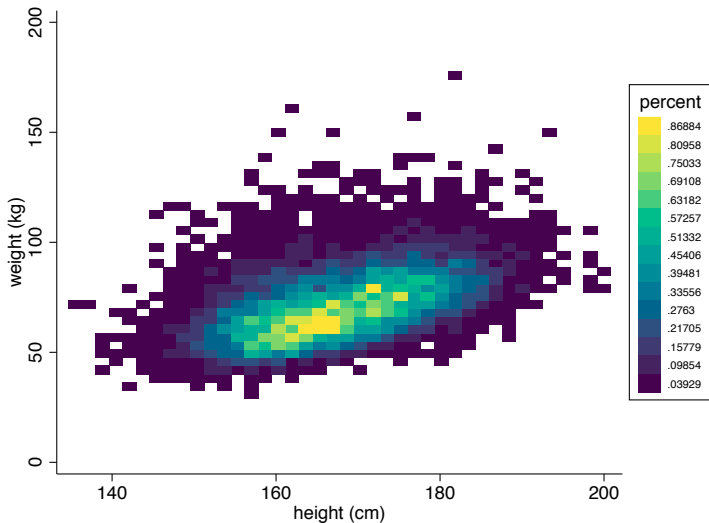
general twoway options

- Some more options related to storing results ...

- 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

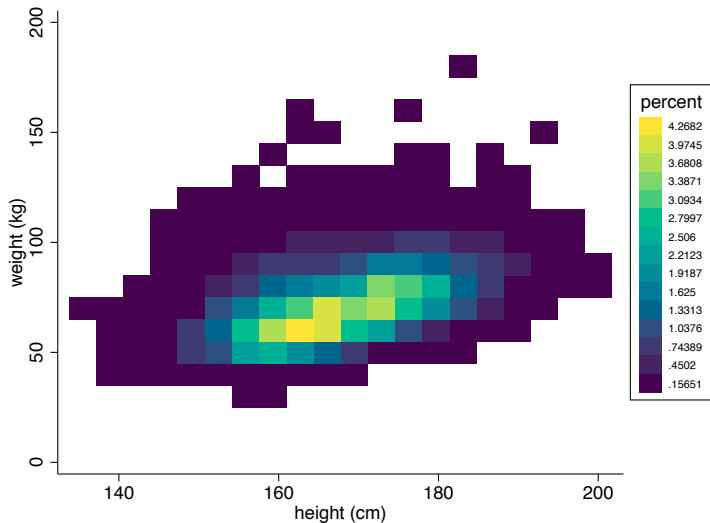
Default

```
. webuse nhanes2, clear  
. heatplot weight height
```



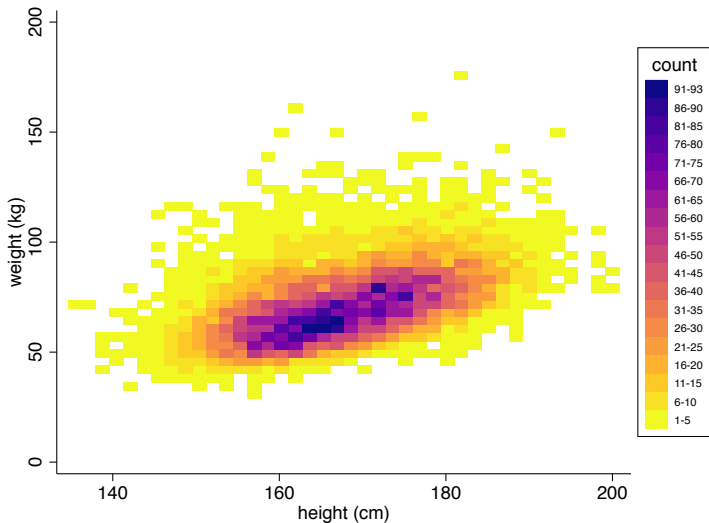
Change resolution

```
. heatmap weight height, xbins(20) ybwidth(10 30)
```



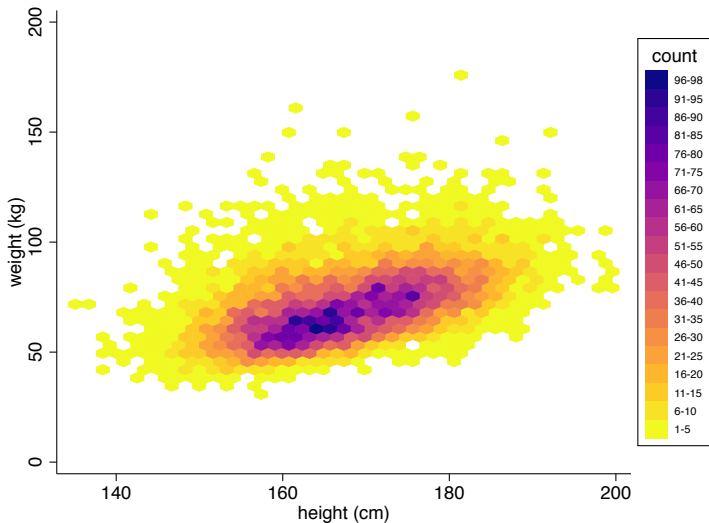
Use counts, change color ramp, change binning, and labeling

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



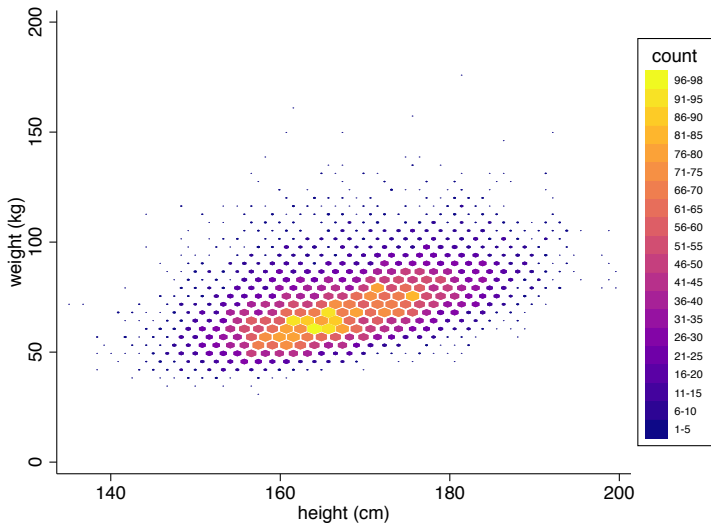
Use hexagons instead of squares

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



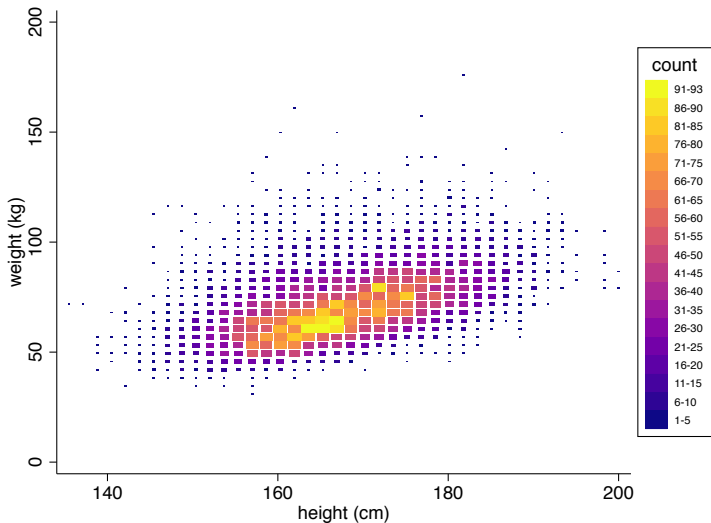
Scale size of hexagons by relative frequency

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



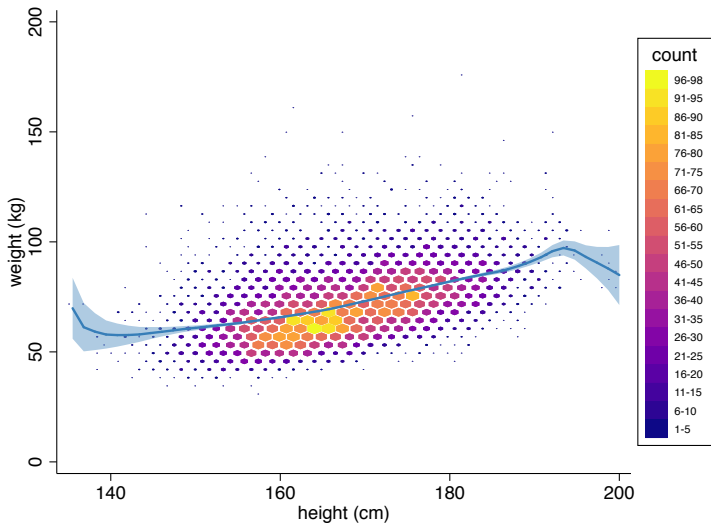
Scaling also available with squares

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



Adding other plots

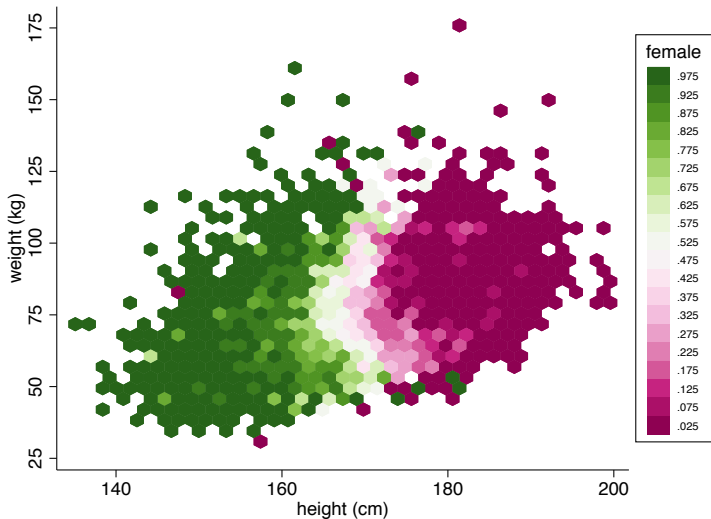
```
. hexplot weight height, statistic(count) color(plasma) ///  
> cut(1(5)@max) keylabels(, range(1)) size ///  
> addplot(lpolyci weight height, degree(1) psty(p2) lw(*1.5) ac(%50) alc(%0))
```



- 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

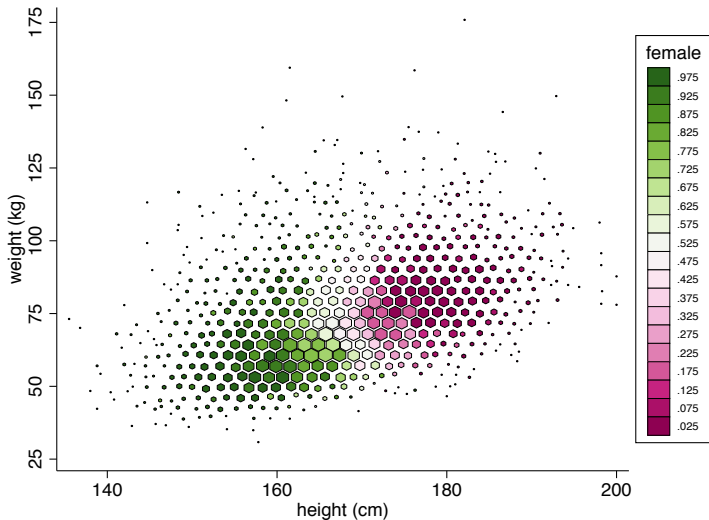
Gender distribution (proportion female) by weight and height

```
. webuse nhanes2, clear  
. hexplot female weight height, color(PiYG) ylabel(25(25)175) cuts(0(.05)1)
```



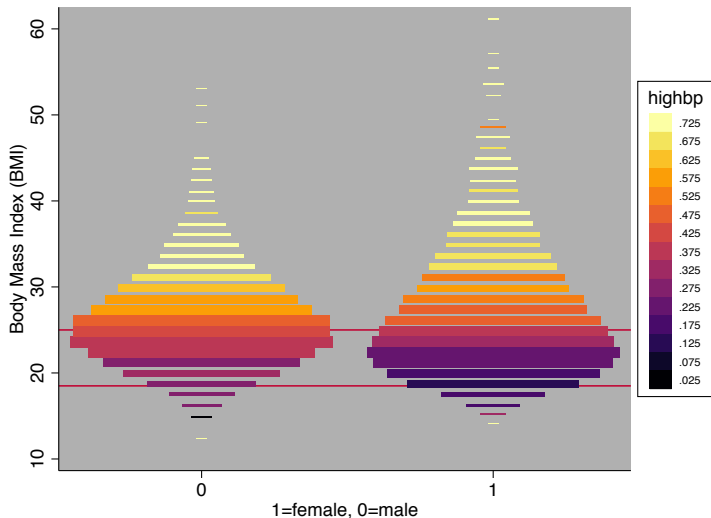
Same graph taking account relative frequencies

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



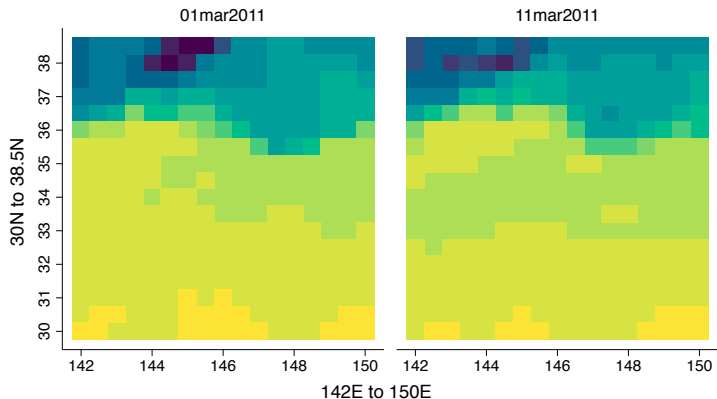
Distribution of the body mass index by gender and its relation to high blood pressure

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



Sea surface temperature by longitude, latitude, and date

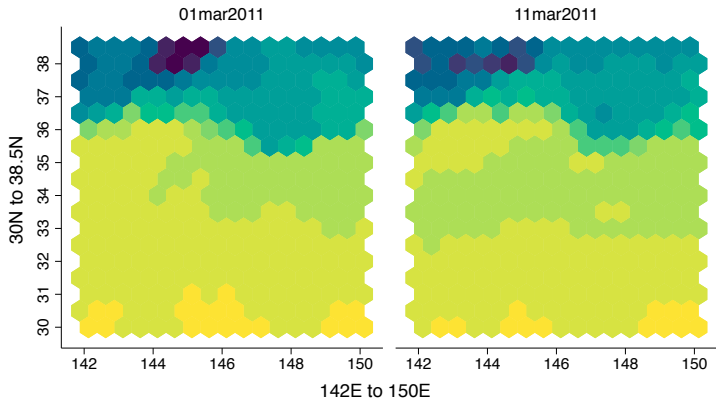
```
. sysuse surface, clear  
(NOAA Sea Surface Temperature)  
. heatplot temperature longitude latitude, discrete(.5) statistic(asis) ///  
> by(date, legend(off)) ylabel(30(1)38) aspectratio(1)
```



Graphs by date

Same plot using hexagons

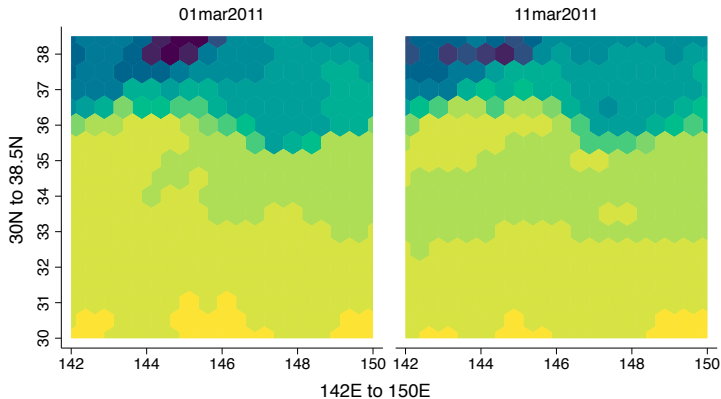
```
. hexplot temperature longitude latitude, discrete(.5) statistic(asis) ///  
> by(date, legend(off)) ylabel(30(1)38) aspectratio(1)
```



Graphs by date

Same plot using hexagons

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

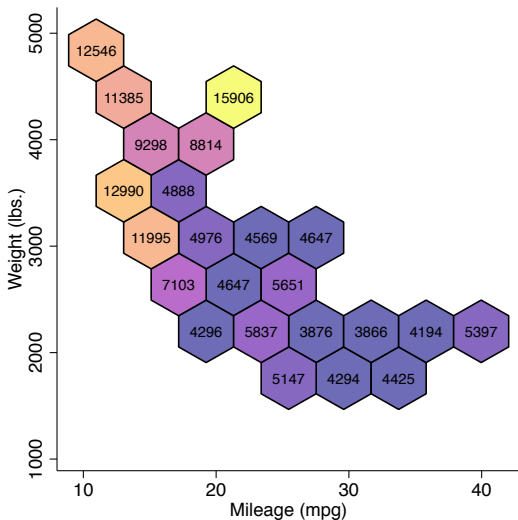


Graphs by date

- 1 Introduction
- 2 Syntax of `heatmap` and `hexplot`
- 3 **Examples**
 - Bivariate histogram
 - Trivariate distributions
 - **Display values as marker labels**
 - Correlation matrix
 - Spatial weights matrix
- 4 Installation

Same plot using hexagons

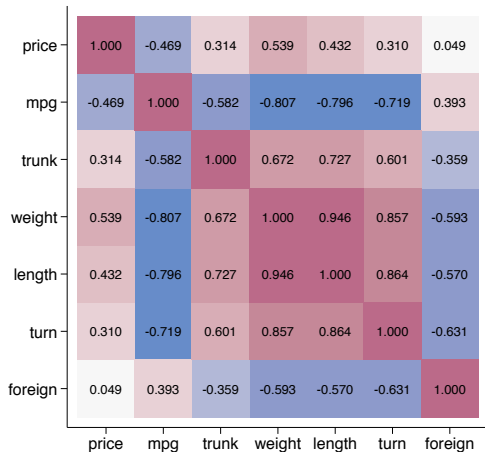
```
. quietly sysuse auto, clear  
. hexplot price weight mpg, values(format(%9.0f)) legend(off) aspectratio(1) ///  
> colors(plasma, intensity(.6)) p(lc(black) lalign(center))
```



- 1 Introduction
- 2 Syntax of `heatmap` and `hexplot`
- 3 **Examples**
 - Bivariate histogram
 - Trivariate distributions
 - Display values as marker labels
 - **Correlation matrix**
 - Spatial weights matrix
- 4 Installation

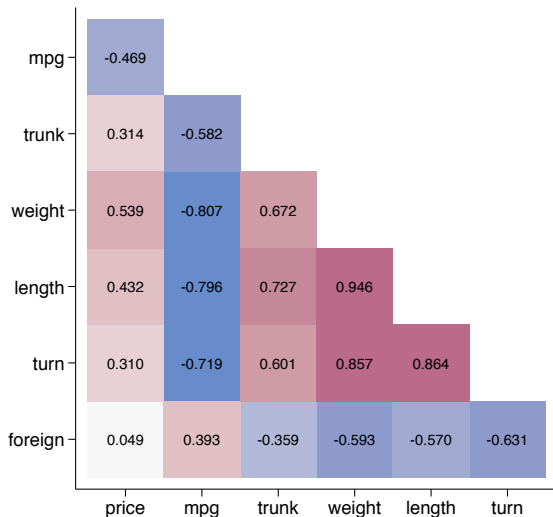
First store correlations in a matrix and then plot from there

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



Plot lower triangle only

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



- 1 Introduction
- 2 Syntax of `heatmap` and `hexplot`
- 3 **Examples**
 - Bivariate histogram
 - Trivariate distributions
 - Display values as marker labels
 - Correlation matrix
 - Spatial weights matrix
- 4 Installation

Copy some data

```
. copy http://www.stata-press.com/data/r15/homicide1990.dta .  
. copy http://www.stata-press.com/data/r15/homicide1990_shp.dta .
```

Compute spacial weights matrix (this might take a while)

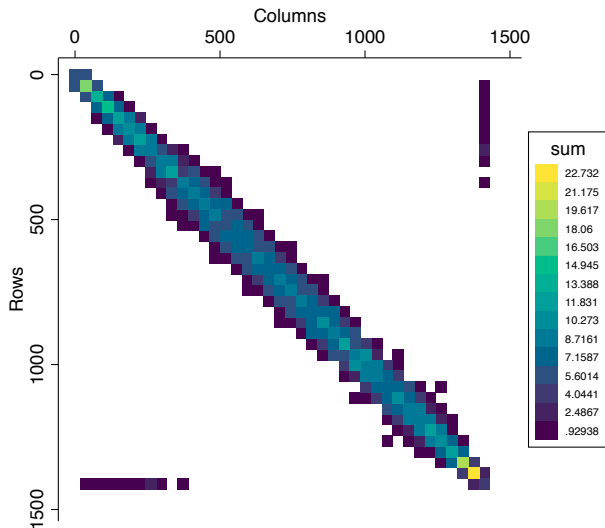
```
. use homicide1990  
(S.Messner et al.(2000), U.S southern county homicide rates in 1990)  
. spmatrix create contiguity W  
. spmatrix matafromsp W id = W  
. mata mata describe W
```

# bytes	type	name and extent
15,949,952	real matrix	W[1412,1412]

(matrix W has about 2 million cells)

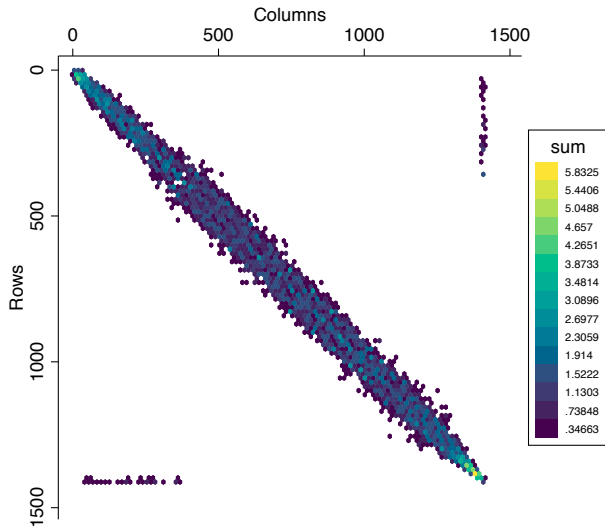
Heat plot of W with default settings, ignoring cells (i.e. weights) that are equal to zero

```
. heatmap mata(W), drop(0) aspectratio(1)
```



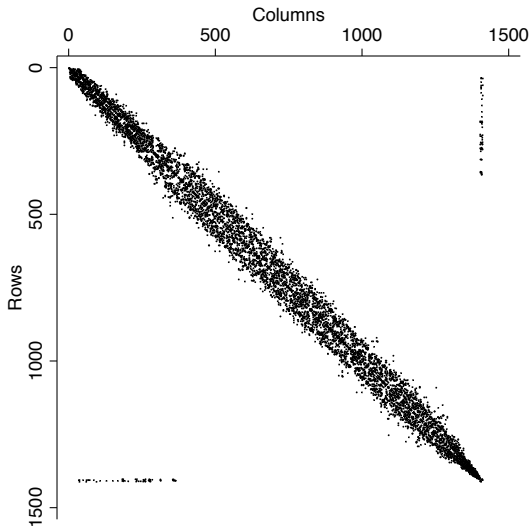
Hexagon plot with fine-grained resolution

```
. heatmap mata(W), drop(0) aspectratio(1) hexagon bins(100)
```



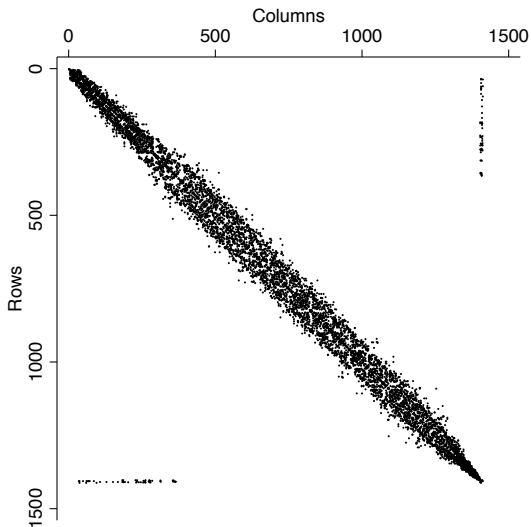
Plot each cell individually using the discrete option

```
. heatmap mata(W), drop(0) aspectratio(1) discrete color(black) p(lalign(center))
```



Could also use the scatter option

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



- 1 Introduction
- 2 Syntax of `heatmap` and `hexplot`
- 3 Examples
 - Bivariate histogram
 - Trivariate distributions
 - Display values as marker labels
 - Correlation matrix
 - Spatial 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
```