

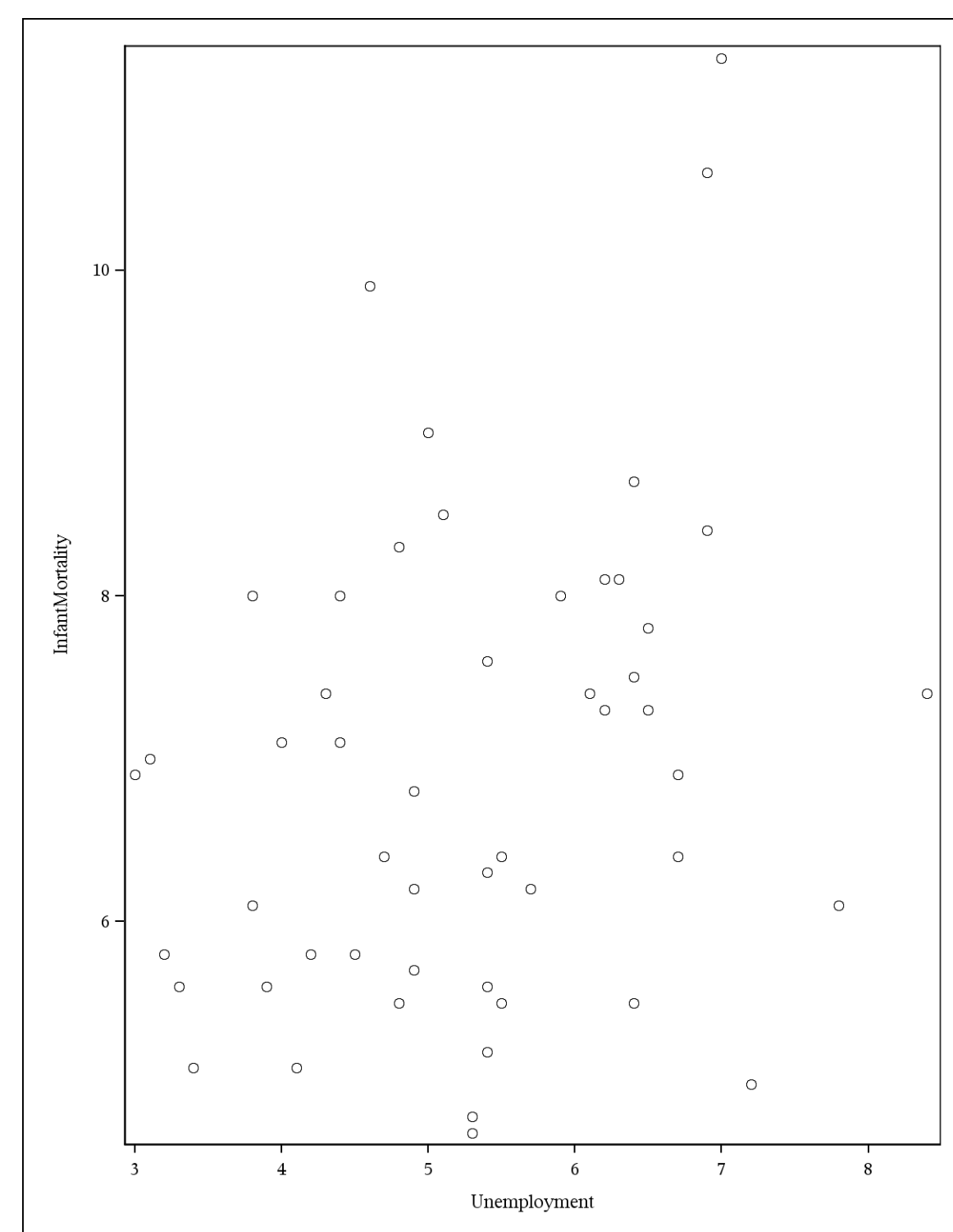
Scatterplots

- Relationship between two continuous variables
- Finding bivariate outliers

Two SAS® tools for scatterplots

- PROC SGPLOT
- Graph Template Language

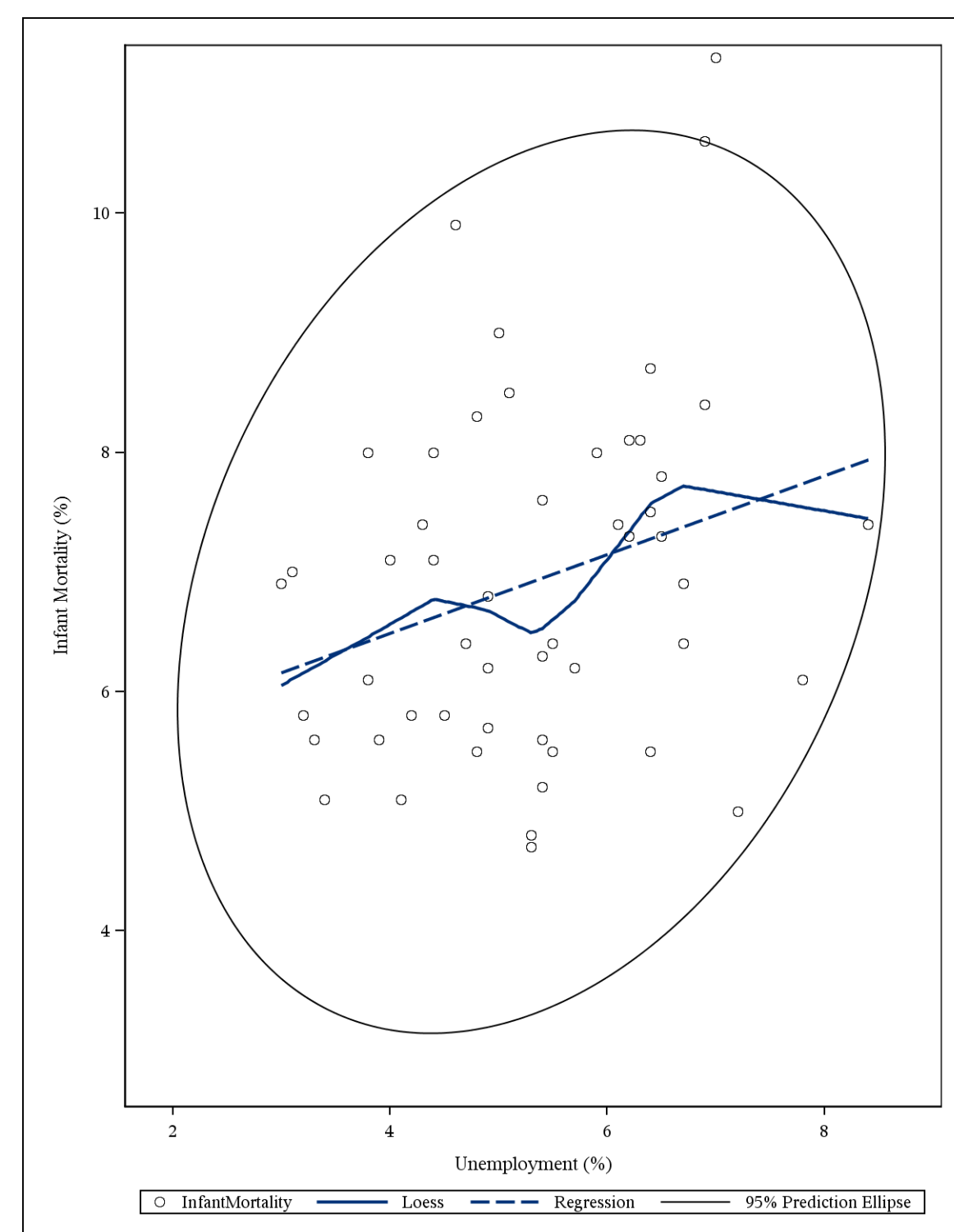
PROC SGPLOT



Many ways to improve it, for example:

- Fix axes
- Add loess smoothed lines
- Add regression line
- Add confidence ellipse

PROC SGPLOT

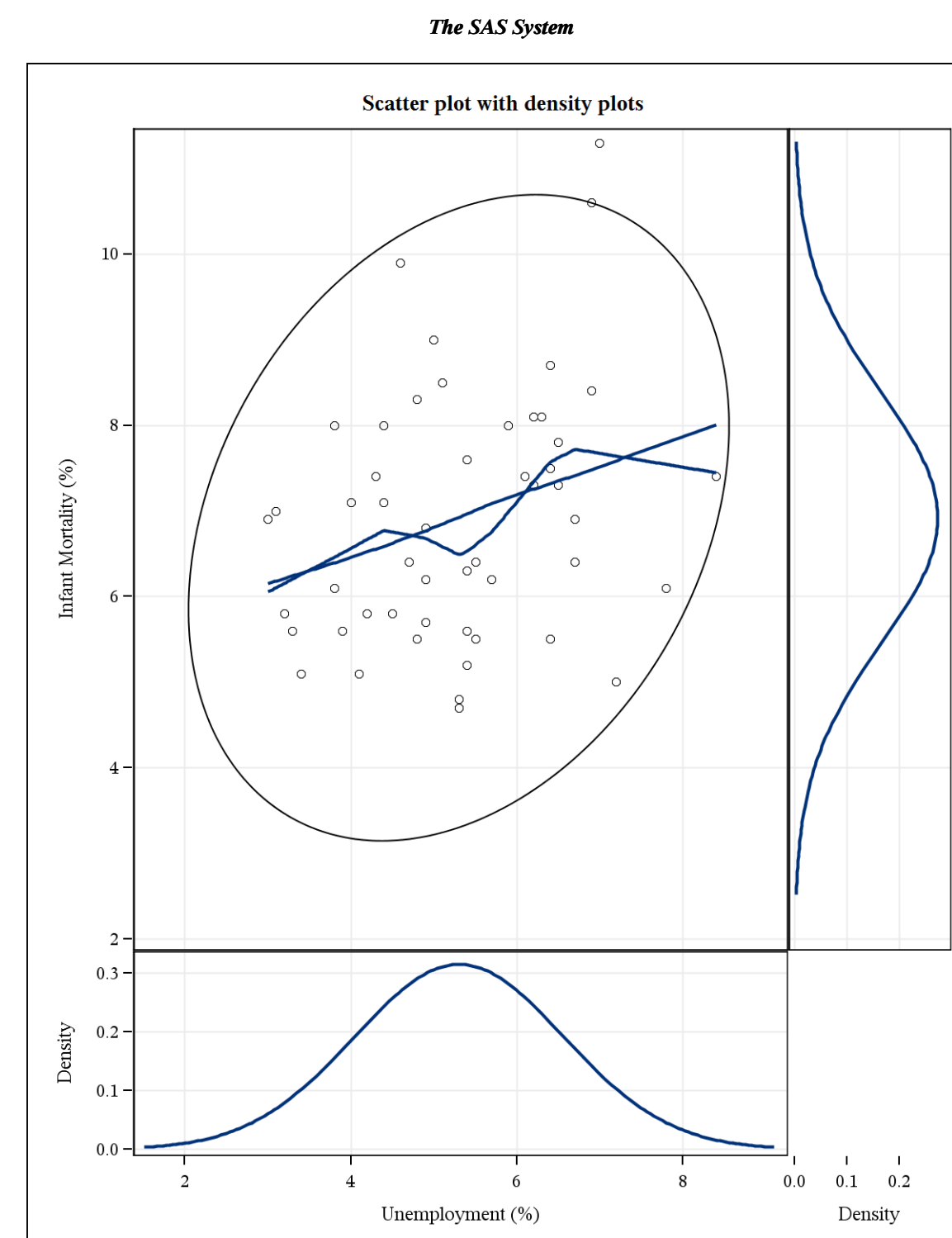


Code Notes:

- PROC SGPLOT
- Scatter $x = y =$
- Loess $x = y = /nomarkers$
- Reg $x = y =$

To visualize the univariate distributions, add density plots

GRAPH TEMPLATE LANGUAGE

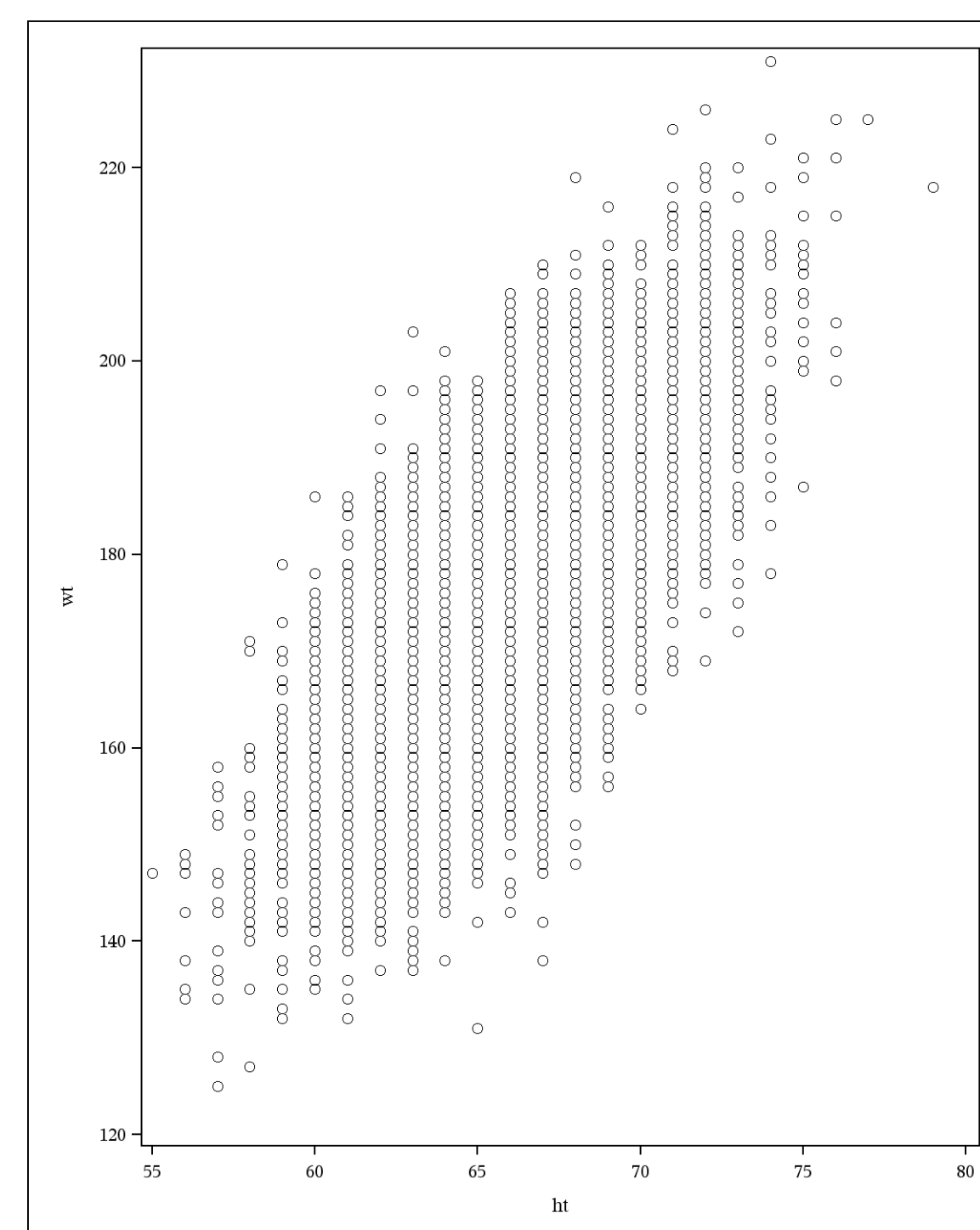


- Create with PROC TEMPLATE
- Render with PROC SGRENDER
- GTL code uses different syntax
- Learning curve; may be worth it

Overplotting is a common problem. Solution depends on size of data set

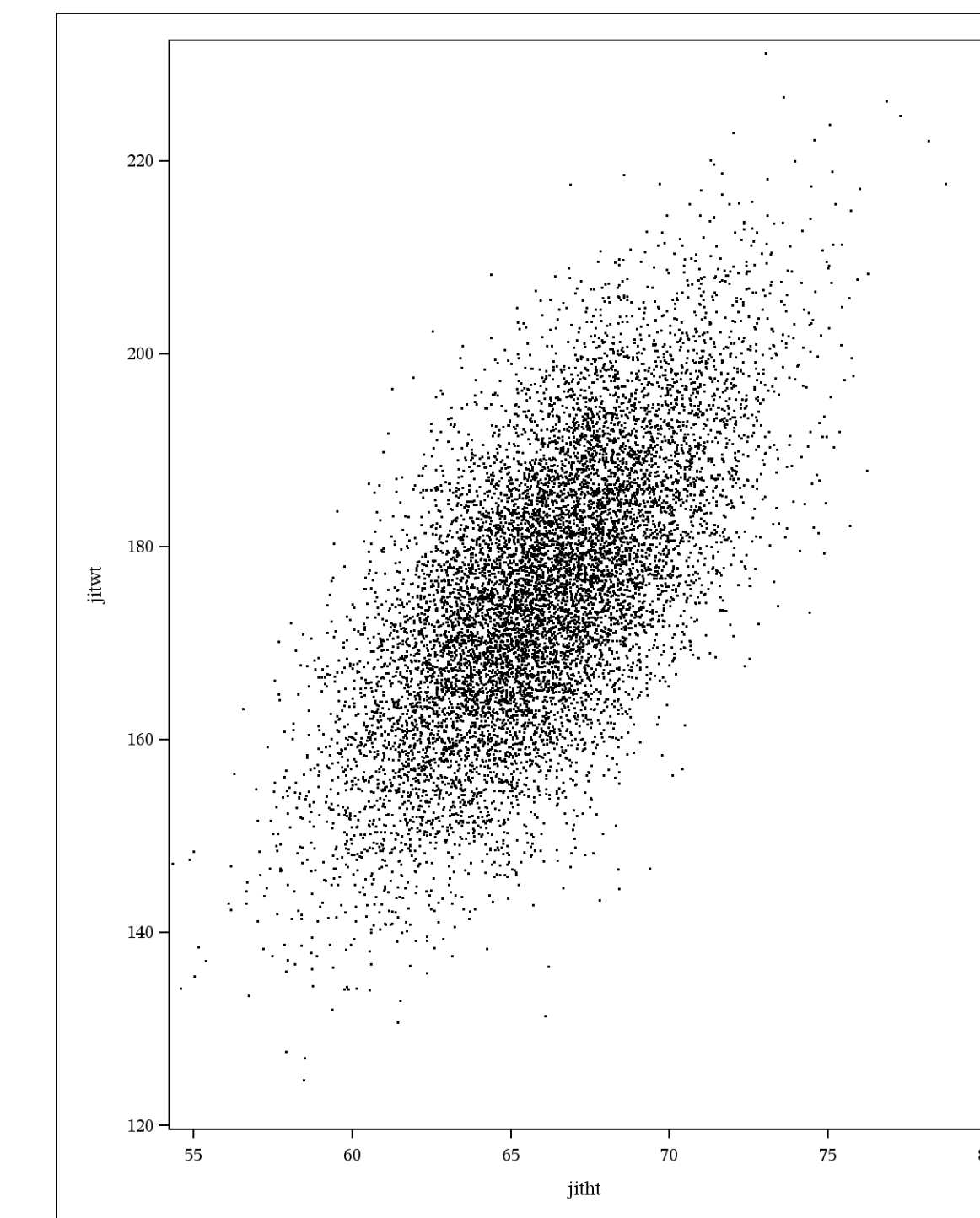
500 heights and weights, rounded to nearest pound or inch

PROC SGPLOT



To solve moderate overplotting use filled circles and make them smaller, jitter rounded data

PROC SGPLOT

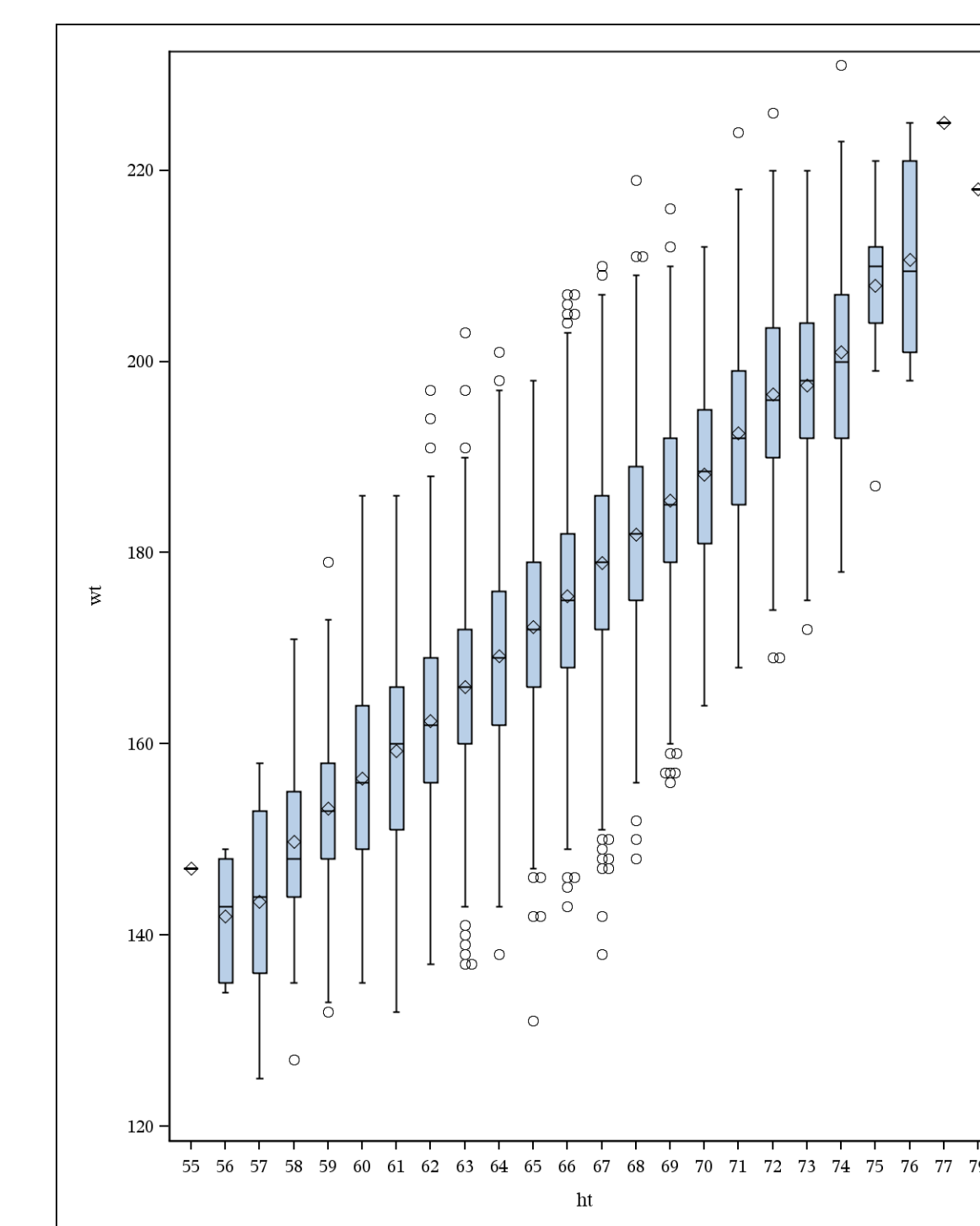


Code Notes:

- Jittering adds noise, must be done in DATA step
- Amount of jitter is judgment call
- MARKERATTRS statement
- SIZE = and SYMBOL = options

To solve extensive overplotting (N = 10,000) categorize X variable and use box plots

PROC SGPLOT



Code Notes:

- PROC SGPLOT
- VBOX wt
- CATEGORY = ht option
- SPREAD option

FULL CODE, WITH NOTES AND MORE EXAMPLES IN PAPER AND HANDOUT