Comments for MEDB 5501, Week 5

Diet data dictionary, 1 of 3

```
data_dictionary: diet.txt

source:
  This data file is part of the Data and
  Story library, an archive of various
  data sets useful for teaching.
  The entire archive is at
  https://dasl.datadescription.com/

description:
  This data set shows side effects of specially
  prepared diet crackers. A more detailed
  description is available at
  https://dasl.datadescription.com/datafile/diet/
```

Speaker notes

Here is the top third of the data dictionary for a file, diet.txt. It comes from the DASL repository.

Diet data dictionary, 2 of 3

```
download:
  https://dasl.datadescription.com/download/data/3163

copyright:
  Unknown. You should be able to use this data for
  individual educational purposes under the Fair Use
  guidelines of U.S. copyright law.

format:
  delimiter: tab
  varnames: first row of data
  missing-value-code: not needed
  rows: 51
  columns: 2
```

Speaker notes

This is a tab delimited file with 51 rows and 2 columns.

Diet data dictionary, 3 of 3

```
vars:
  Bloat:
    label: Did the patient experience bloating?
    format: string

  Cracker:
    label: Type of cracker
    format: string
```

Speaker notes

There are two variables, Bloat and Cracker. Both are strings.

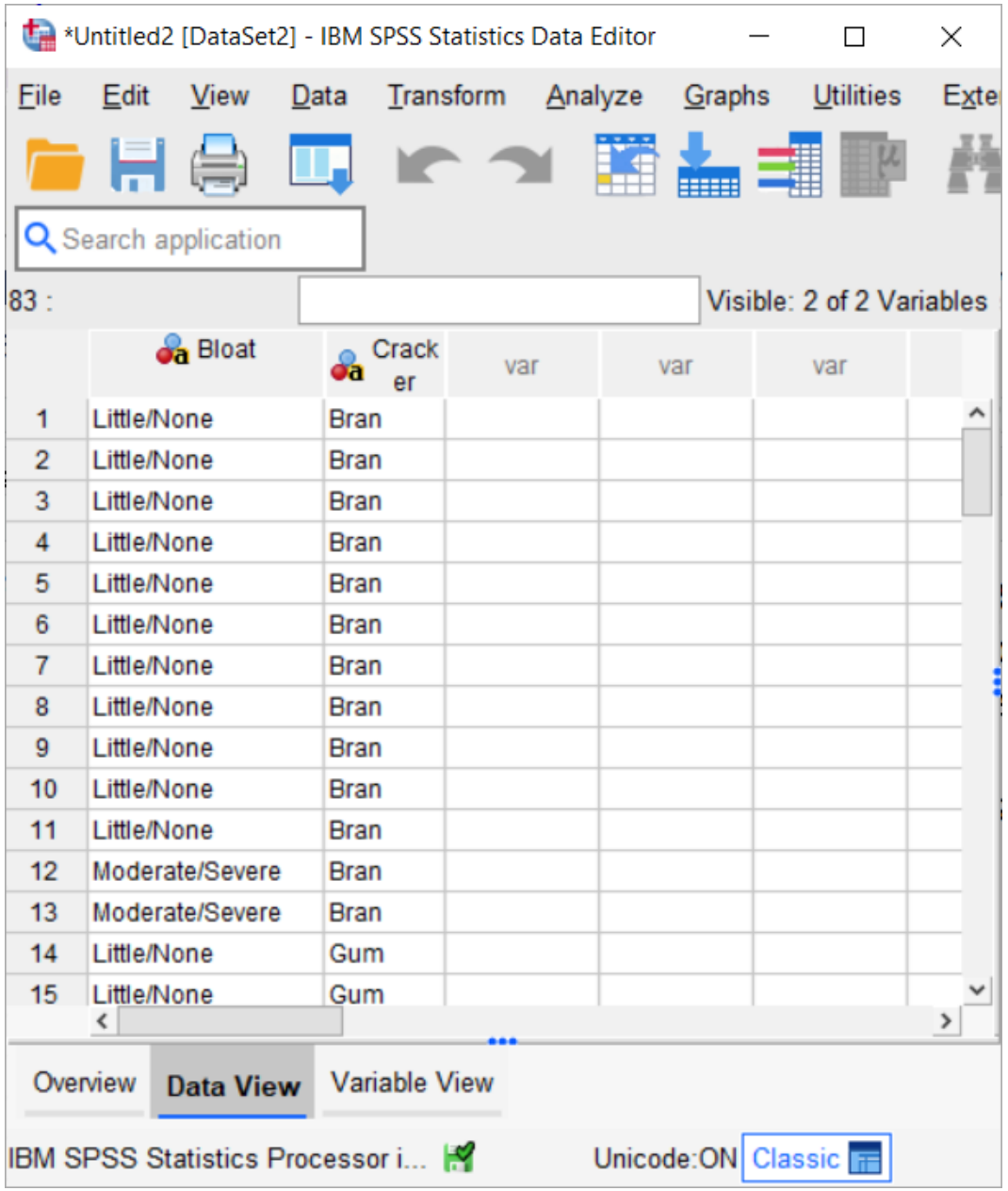

Figure 1: SPSS Data View

Speaker notes

Here is what the first few rows of data show.

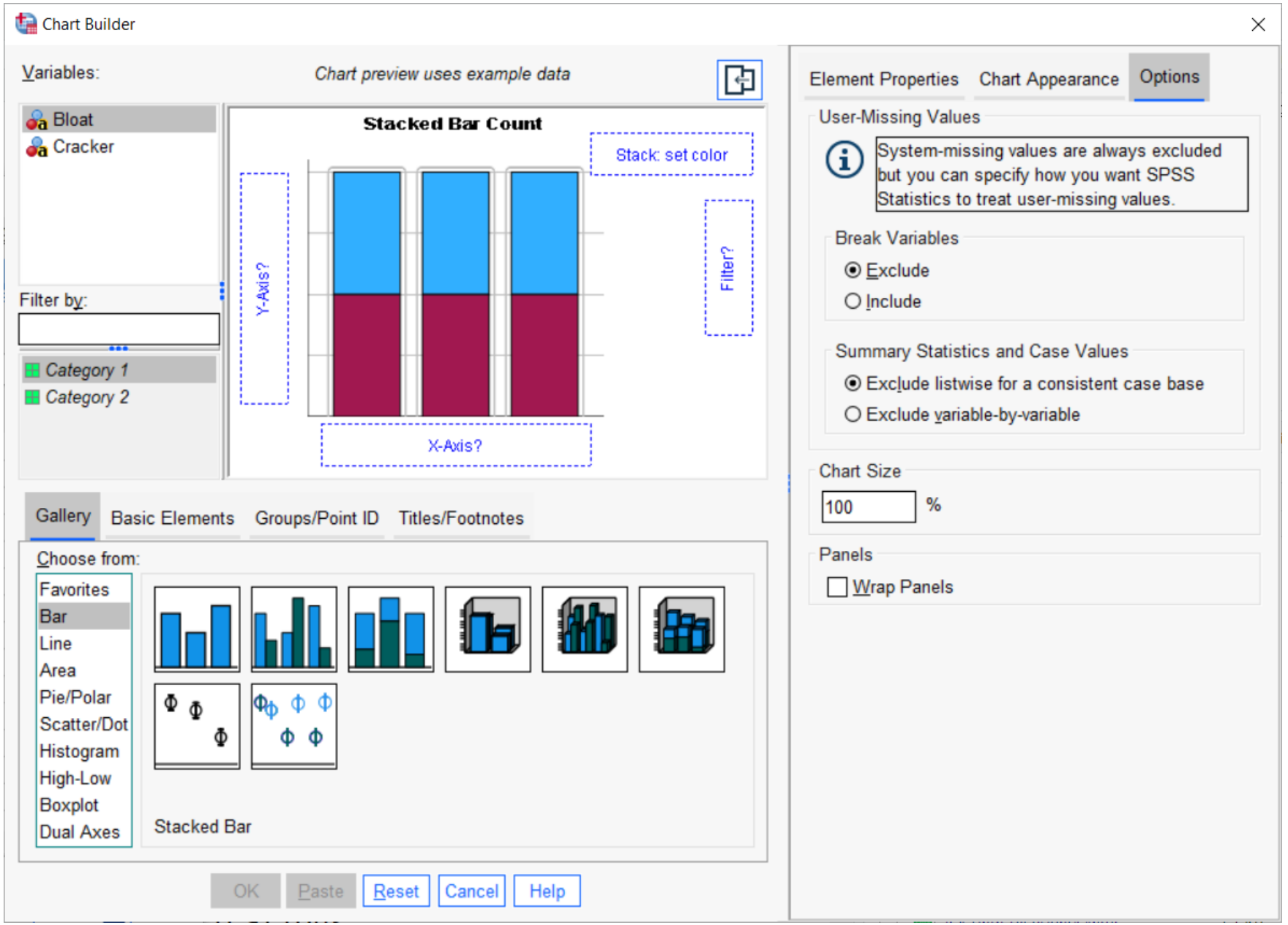

Figure 2: SPSS Chart Builder, stacked bar chart, before adding variables

Speaker notes

To create any type of chart, select Graphs | Chart Builder from the SPSS menus. Then drag and drop the chart type. Here, you see the stacked bar chart, represented by the icon in the first row, third from the left.

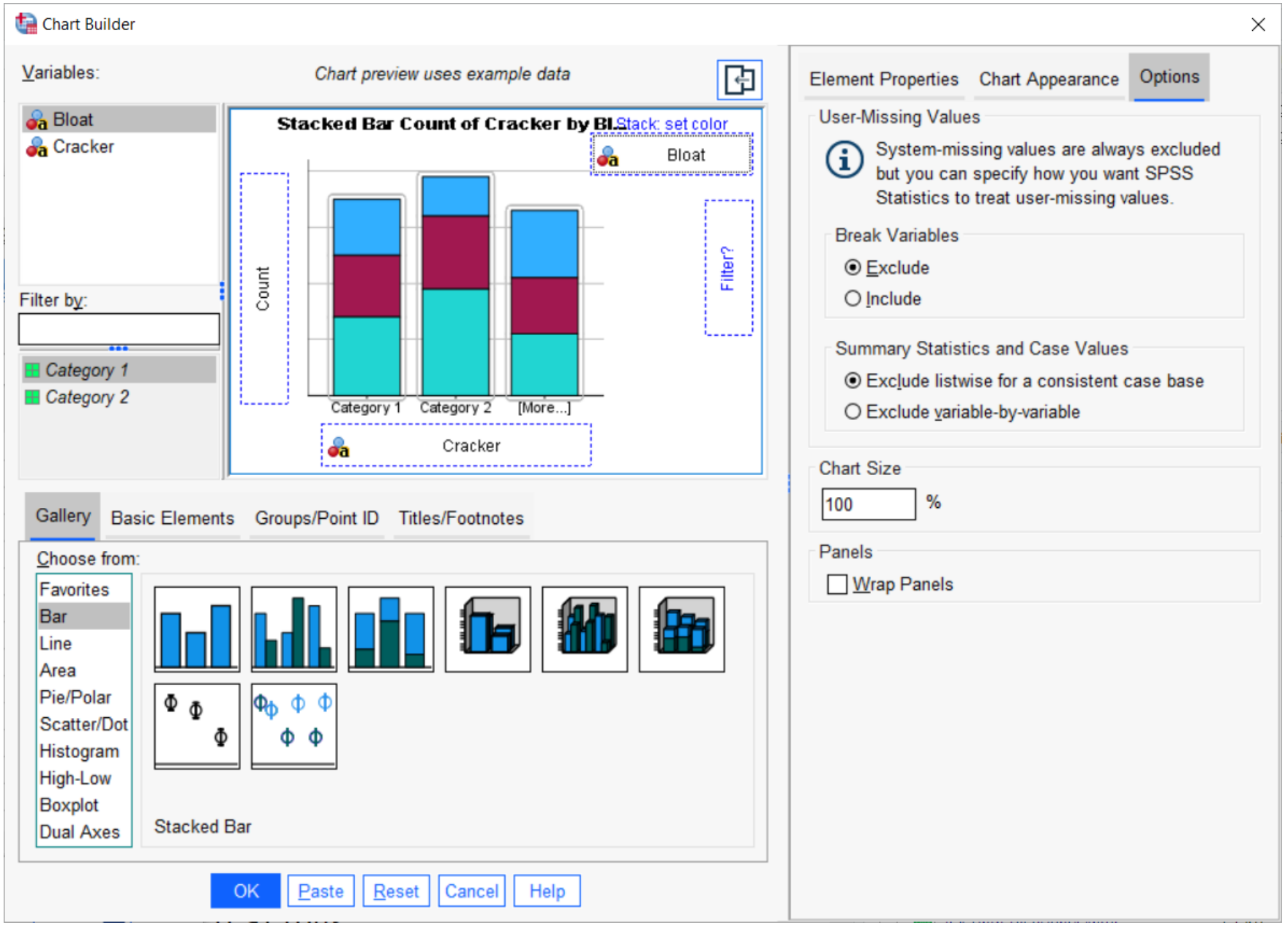

Figure 3: SPSS Chart Builder, stacked bar chart, after adding variables

Speaker notes

Drag and drop the variables into the X-axis and Stack windows.

GGraph

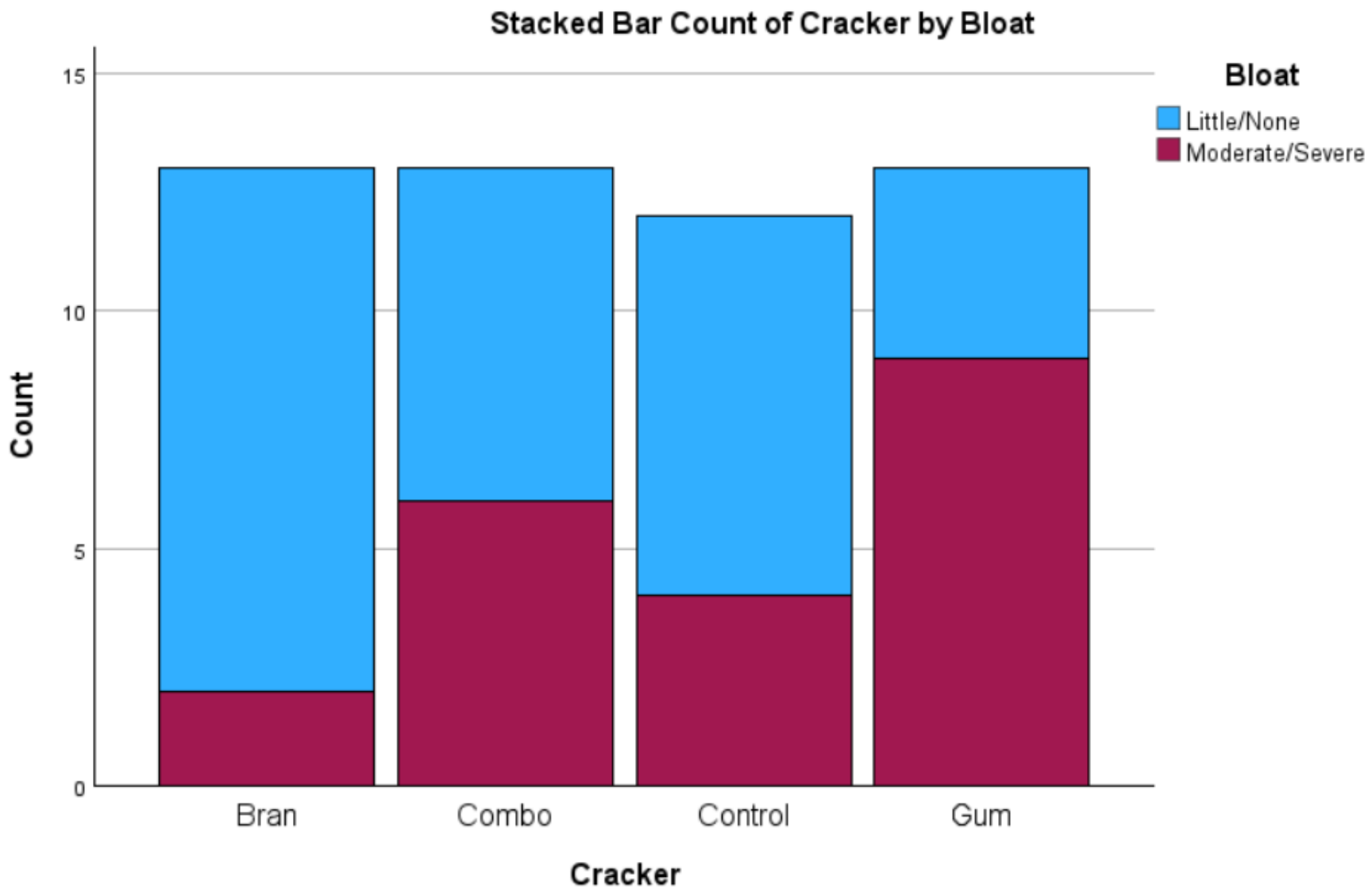

Figure 4: SPSS output, original (default) bar chart

Speaker notes

Here is the default graph produced by SPSS. Let’s see what modifications can be made.

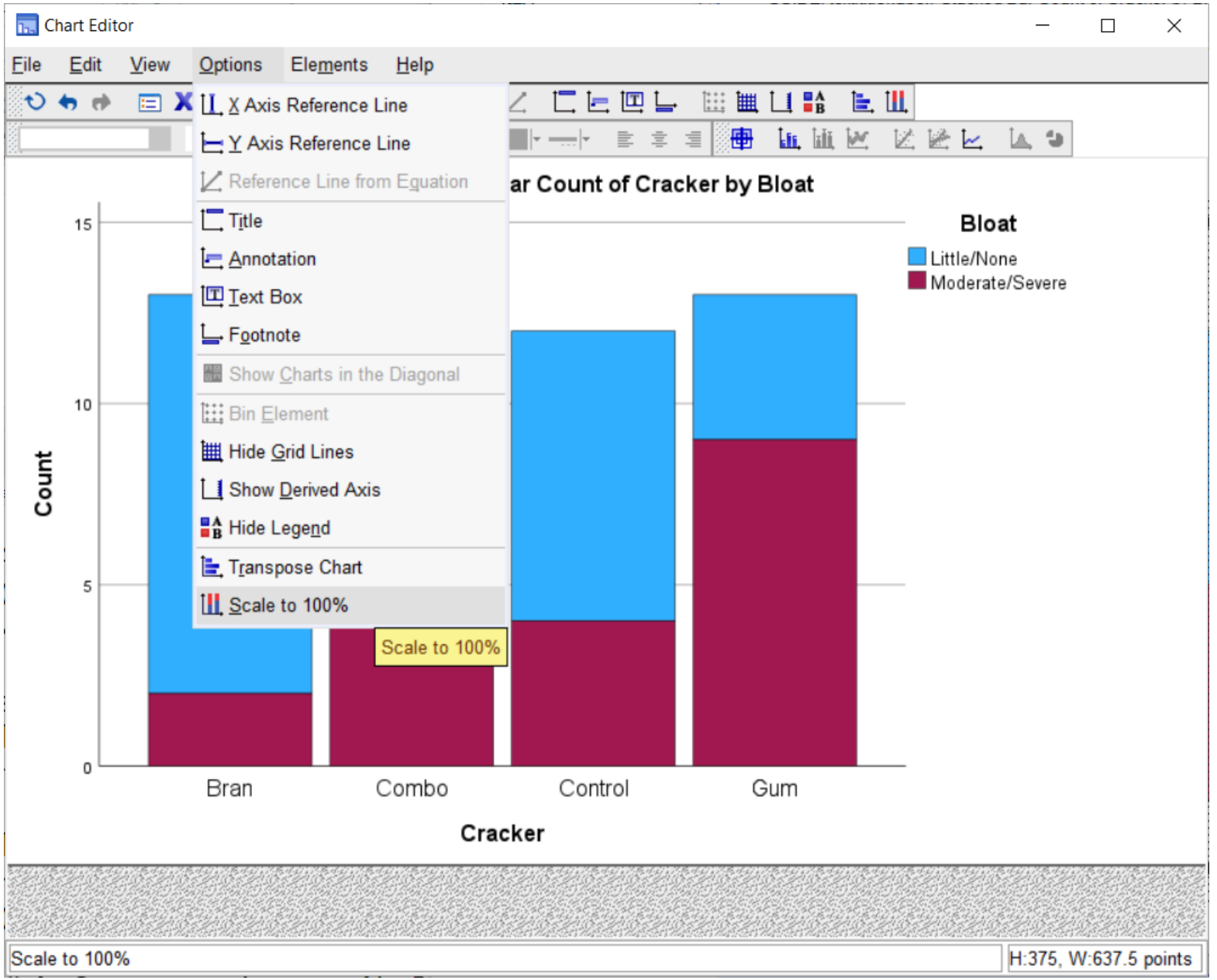

Figure 5: SPSS Properties dialog box, options menu

Speaker notes

Select Options | Scale to 100% to use percentages rather than counts.

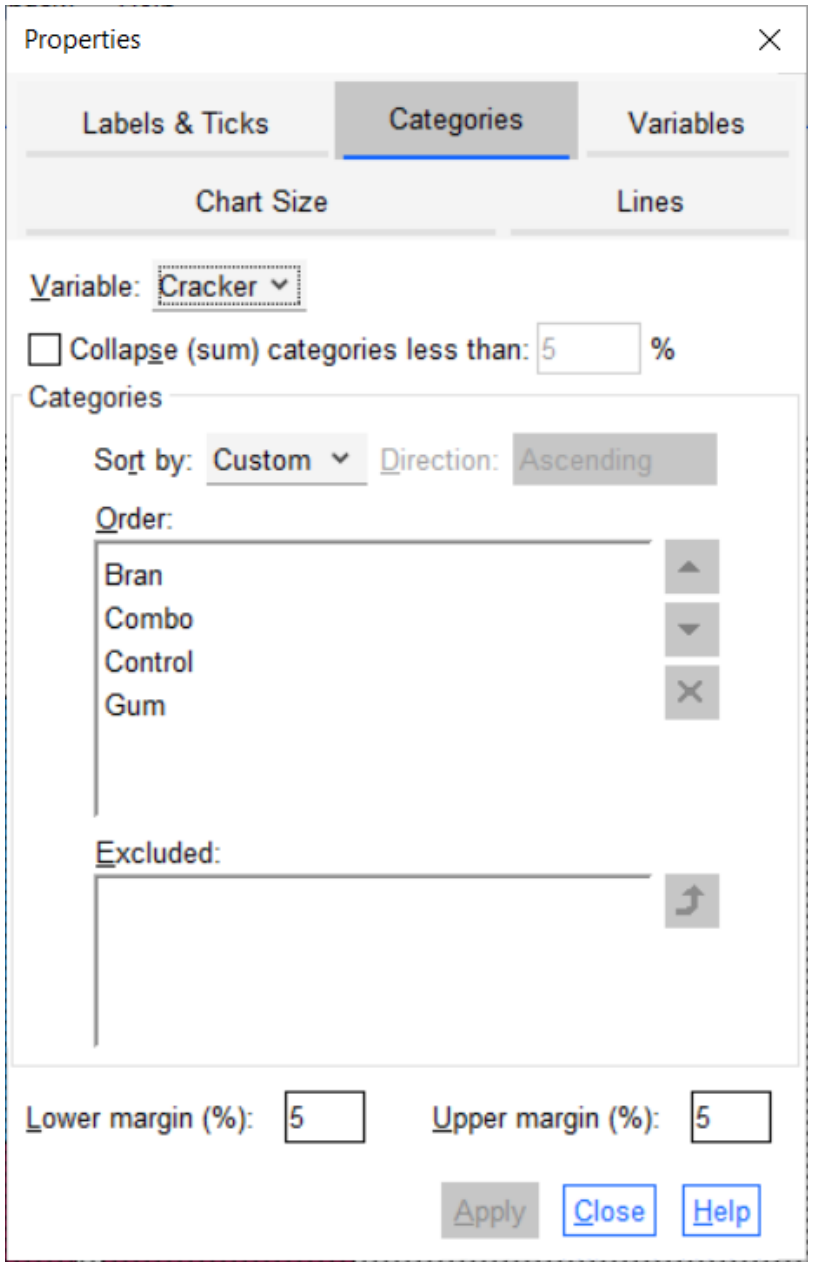

Figure 6: SPSS Properties dialog box, Categories tab

Speaker notes

This dialog box allows you to change the order of the bars.

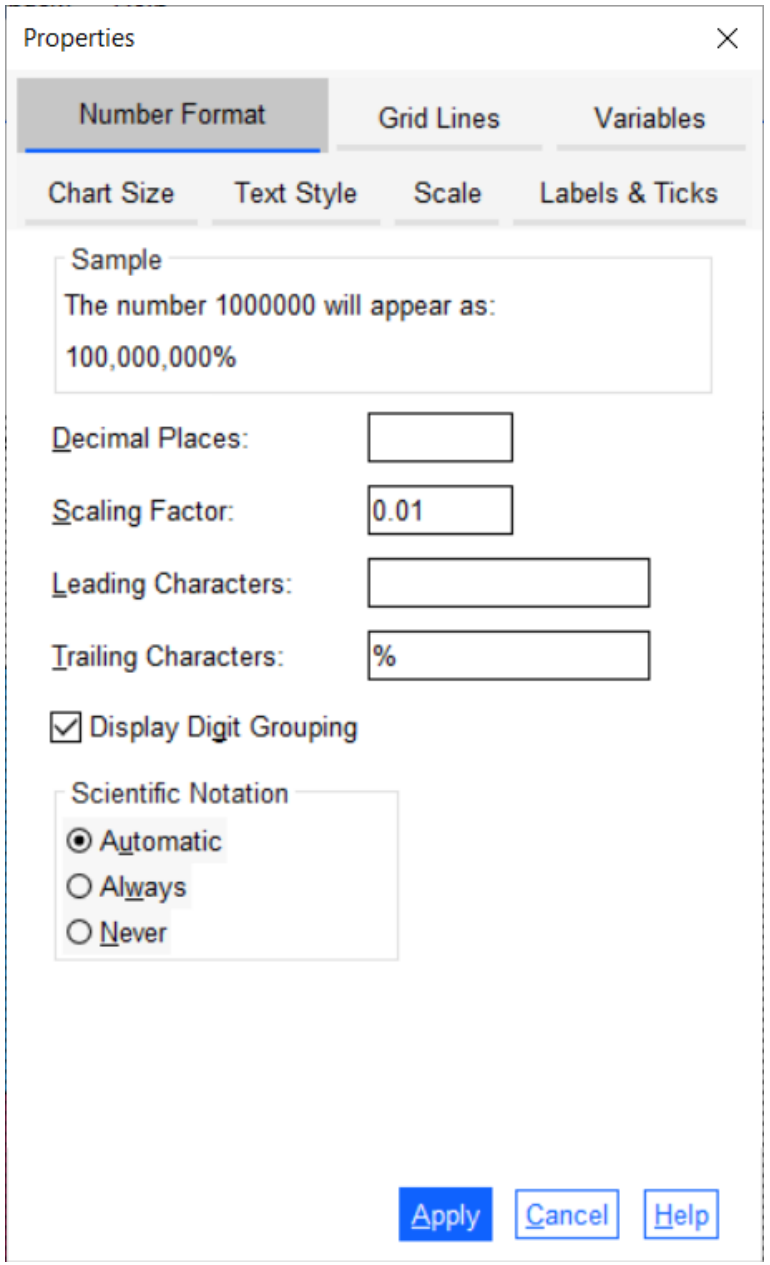

Figure 7: SPSS Properties dialog box, Number Format tab

Speaker notes

This dialog box allows you to control the number of decimals displayed on the axes.

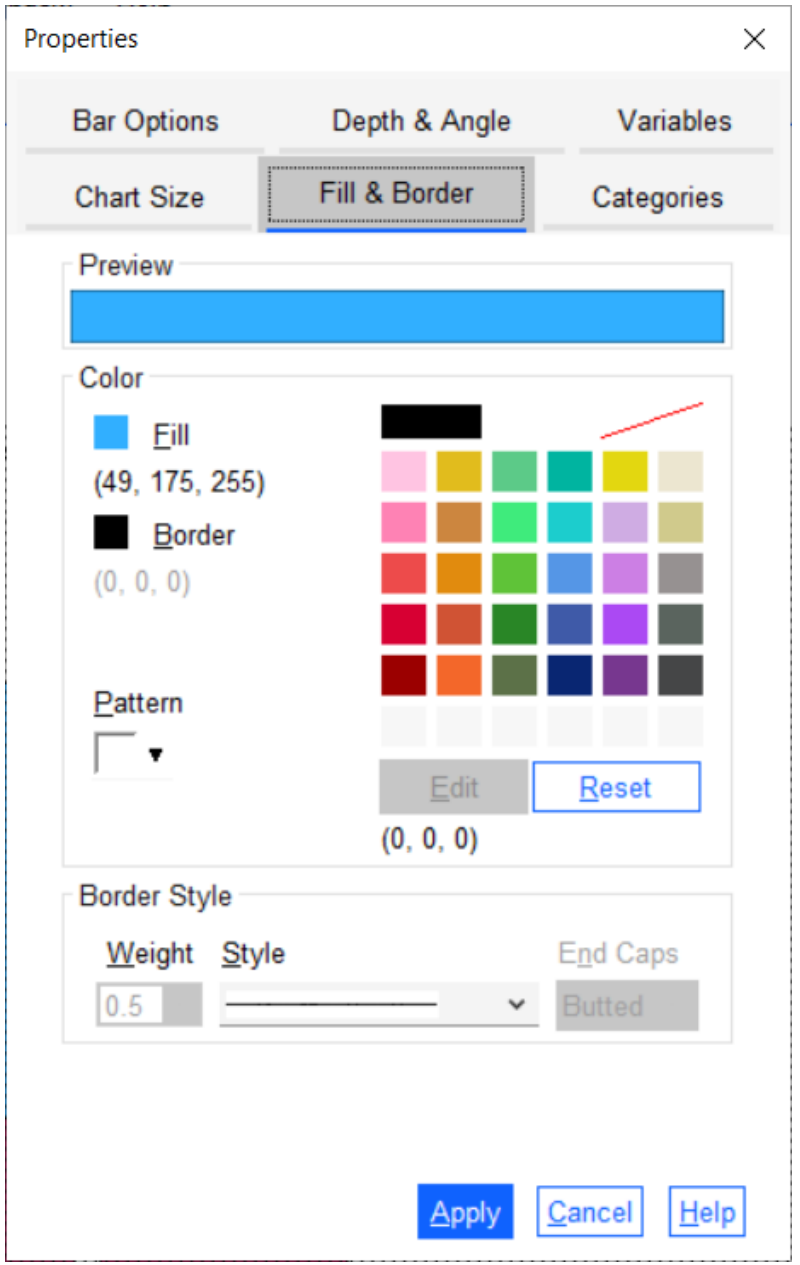

Figure 8: SPSS Properties dialog box, Fill & Border tab

Speaker notes

This dialog box allows you to change the fill and border colors of the bars.

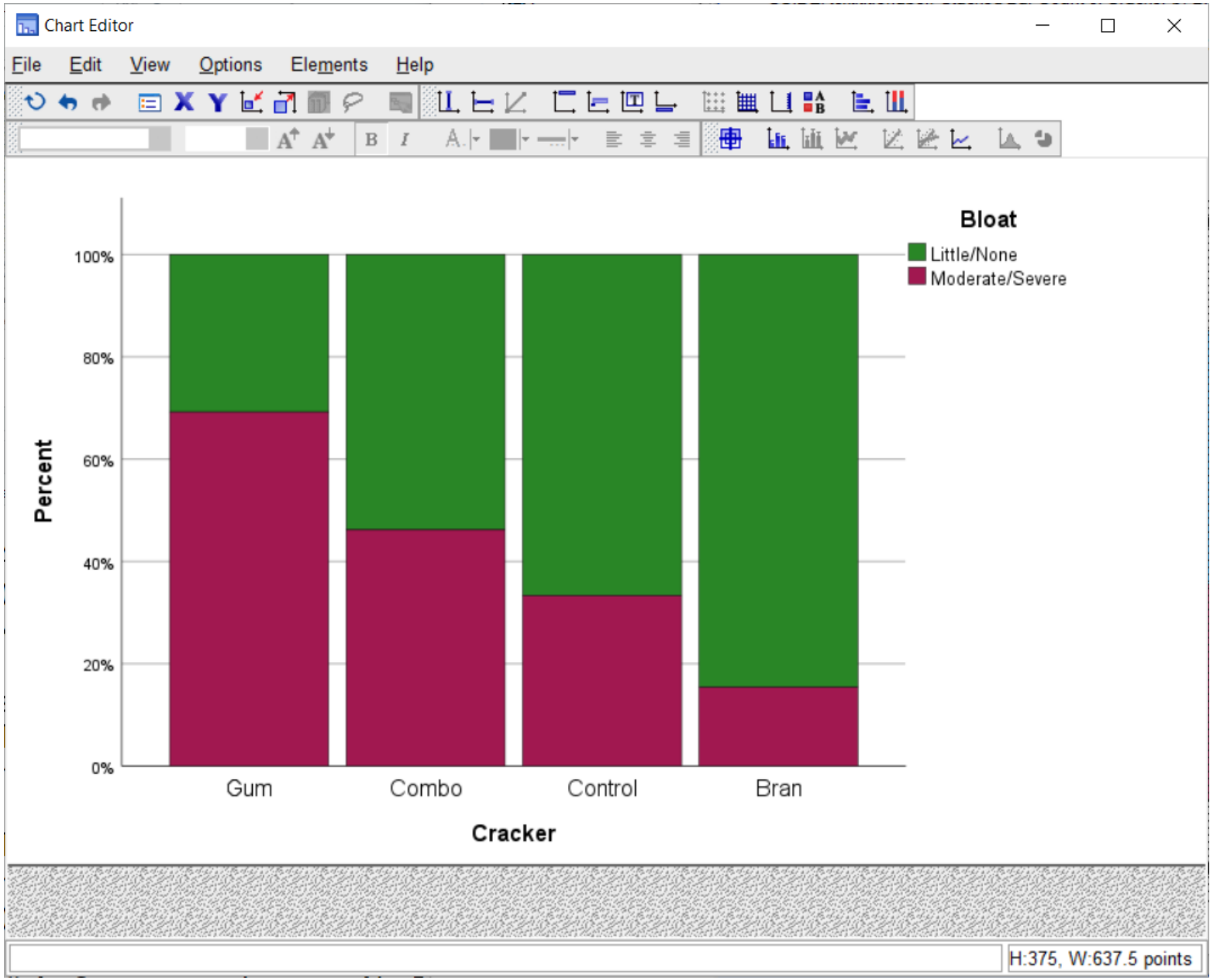

Figure 9: SPSS output, modified bar chart

Speaker notes

Here is the modified bar chart.

Break 1

- What have you learned?
  - Creating and modifying stacked bar chart
- What’s coming next
  - Titanic data, stacked bar chart

Speaker notes

Time for a break. Any questions?

Titanic data dictionary, 1 of 3

```
data_dictionary: titanic.txt

description:
  Mortality among passengers of the Titanic
  http://www.statsci.org/data/general/titanic.html

download:
  http://www.statsci.org/data/general/titanic.txt
  http://www.pmean.com/15/images/day2titanic.txt
```

Speaker notes

Ths is the data dictionary for the titanic data. You can find a nice description of the data at the website listed here.

Titanic data dictionary, 2 of 3

```
vars:
  Name:
    label: Passenger name

  PClass:
    label: Passenger class
    scale: ordinal text categories
    values: 1st, 2nd, 3rd

  Age:
    unit: years
    scale: positive real
    missing: NA
```

Speaker notes

This shows details for the first three variables in the titanic data set.

Titanic data dictionary, 3 of 3

```
Sex:
  scale: nominal text categories
  values: female, male

Survived:
  scale: binary integer categories
  values:
    yes: 1
    no: 0
```

Speaker notes

This is information on the remaining two variables.

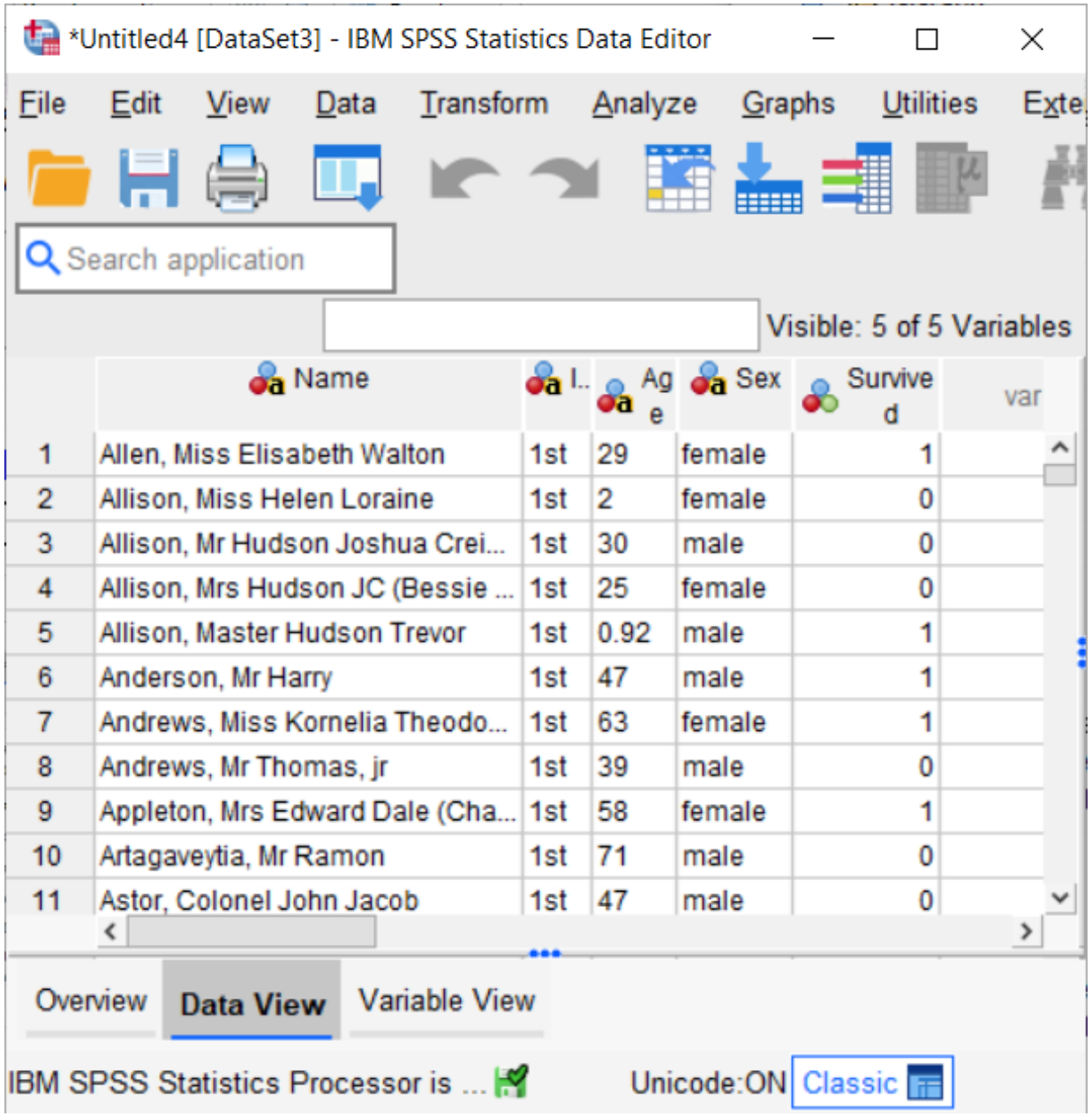

Figure 10: SPSS Data View

Speaker notes

This is a listing of the first few rows of the data.

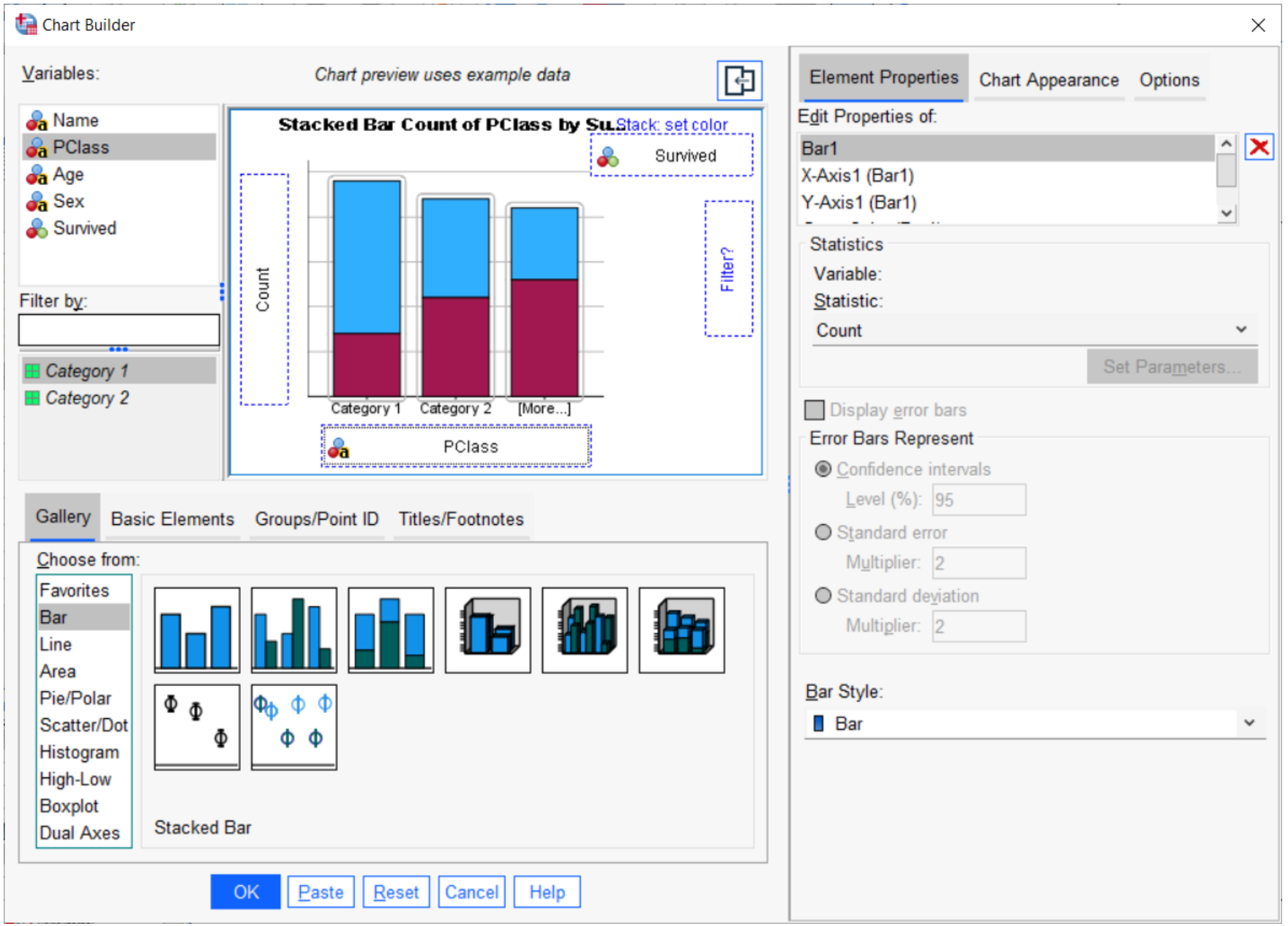

Figure 11: SPSS Chart Builder, stacked bar chart

Speaker notes

To create a stacked bar chart, select Graph | Chart Builder from the SPSS menu. Drag and drop the stacked bar chart icon (first row, third from the left). Drag and drop one categorical variable in the x-axis box and one in the stack box.

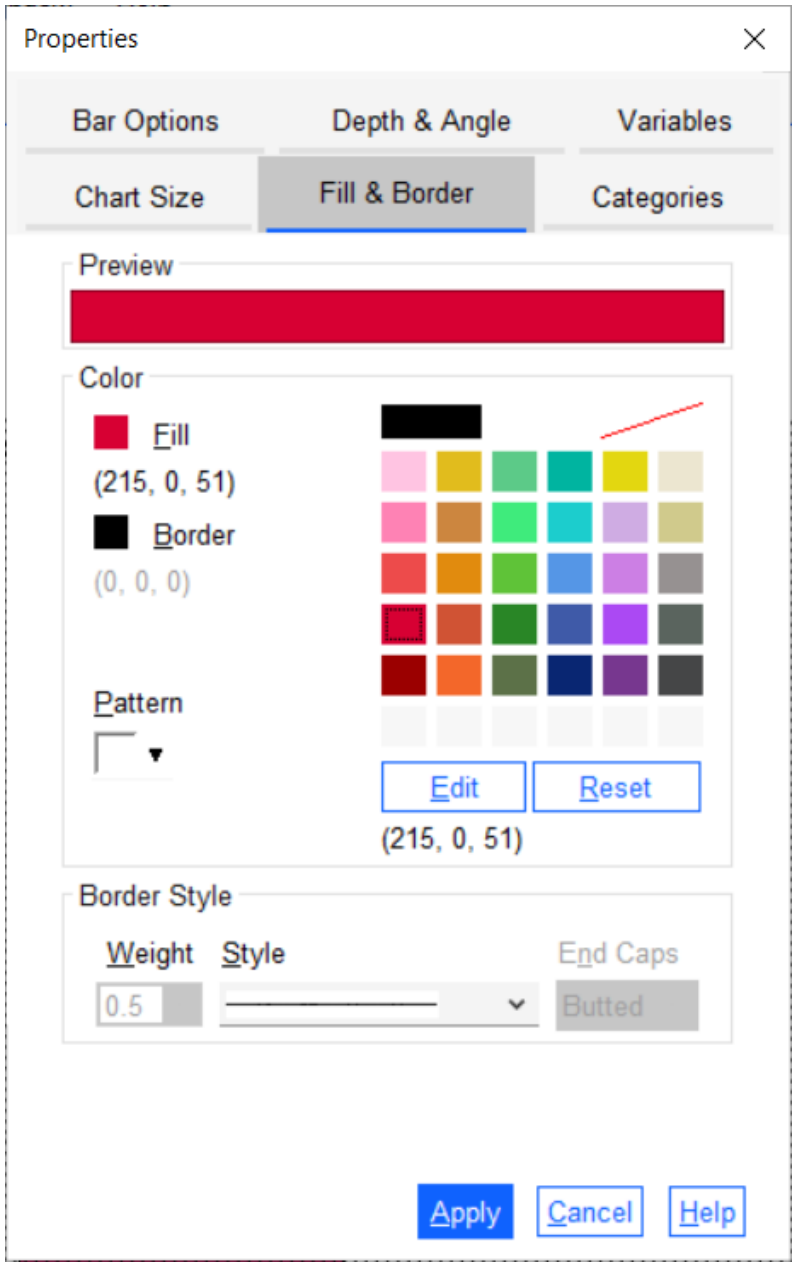

Figure 12: SPSS Properties dialog box, Fill & Border tab

Speaker notes

If you wish to modify the bar chart, double click on the graph. This calls up the properties dialog box. The properties dialog box has different tabs to allow you to make specific changes to individual features of the bar chart. Click on the middle of one of the bars and you should see a change in the dialog box. Click on the Fill & Border tab. You can now change the color of the bar.

Notice the two choices, white and transparent, at the top of the color choices. White will hide anything behind it. Transparent will not.

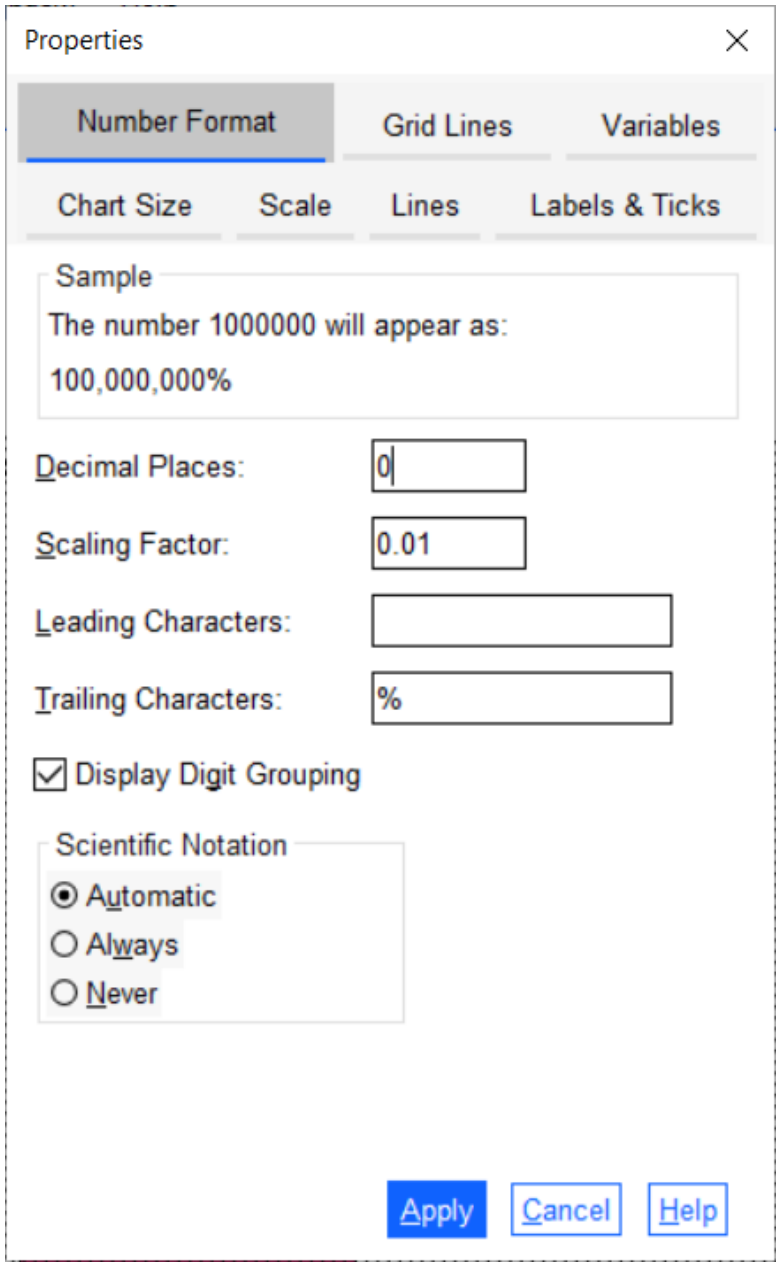

Figure 13: SPSS Properties Dialog Box, Number Format tab

Speaker notes

Add note.

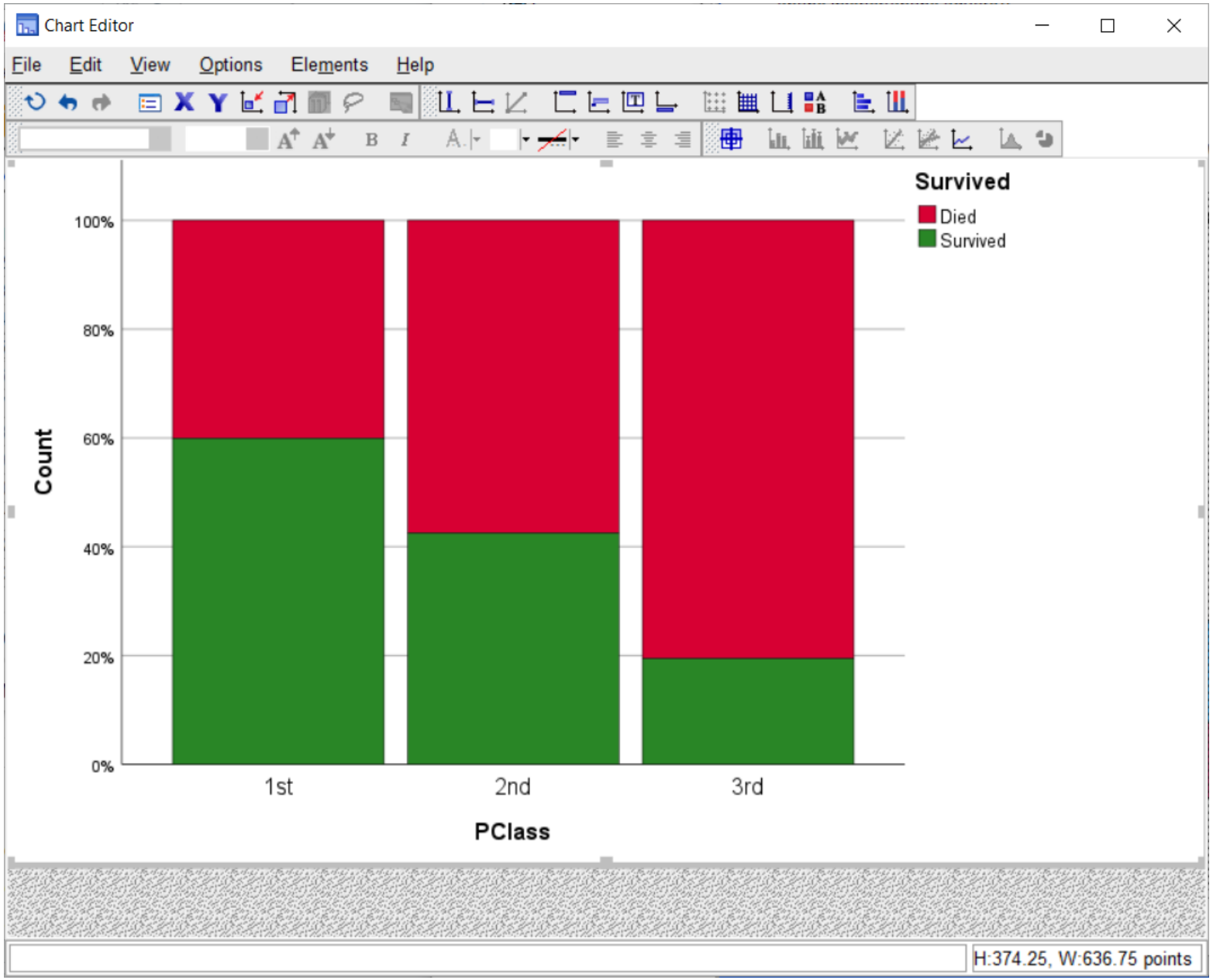

Figure 14: SPSS output, modified bar chart

Speaker notes

Add note.

Break 2

- What have you learned?
  - Titanic data, stacked bar chart
- What’s coming next
  - Housing data, scatterplot

Speaker notes

Time for a break. Any questions?

Housing data dictionary, 1

```
source:
  This file was found originally at a website
  DASL (Data And Story Library) that is no
  longer available.

description:
  The original source describes the data as
  "a random sample of records of resales of
  homes from Feb 15 to Apr 30, 1993 from the
  files maintained by the Albuquerque Board
  of Realtors. This type of data is
  collected by multiple listing agencies in
  many cities and is used by realtors as an
  information base."
```

Speaker notes

Add note.

Housing data dictionary, 2

```
copyright:
    Unknown. You should be able to use this data for
  individual educational purposes under the Fair Use
  guidelines of U.S. copyright law.

format:
  delimiter: space
  varnames: first row of data
  missing-value-code: *
  rows: 117
  columns: 8
```

Speaker notes

Add note.

Housing data dictionary, 3

```
vars:
  Price:
    label: Selling price
    unit: dollars

  SquareFeet:
    label: Living space
    unit: square feet

  AgeYears:
    label: Age of home
    unit: years
```

Speaker notes

Add note.

Housing data dictionary, 4

```
NumberFeatures:
  label:
    Home features (dishwasher, refrigerator,
    microwave, disposer, washer, intercom,
    skylight(s), compactor, dryer, handicap
    fit, cable TV access)
  scale: count
  range: 0 to 11

Northeast:
  label: Located in northeast sector of city?
  values:
    Yes: 1
    No: 0
```

Speaker notes

Add note.

Housing data dictionary, 5

```
CustomBuild:
  label: Custom built?
  values:
    Yes: 1
    No: 0

CornerLot:
  label: Corner location?
  values:
    Yes: 1
    No: 0

Tax:
  label: Yearly property tax
  unit: dollars
```

Speaker notes

Add note.

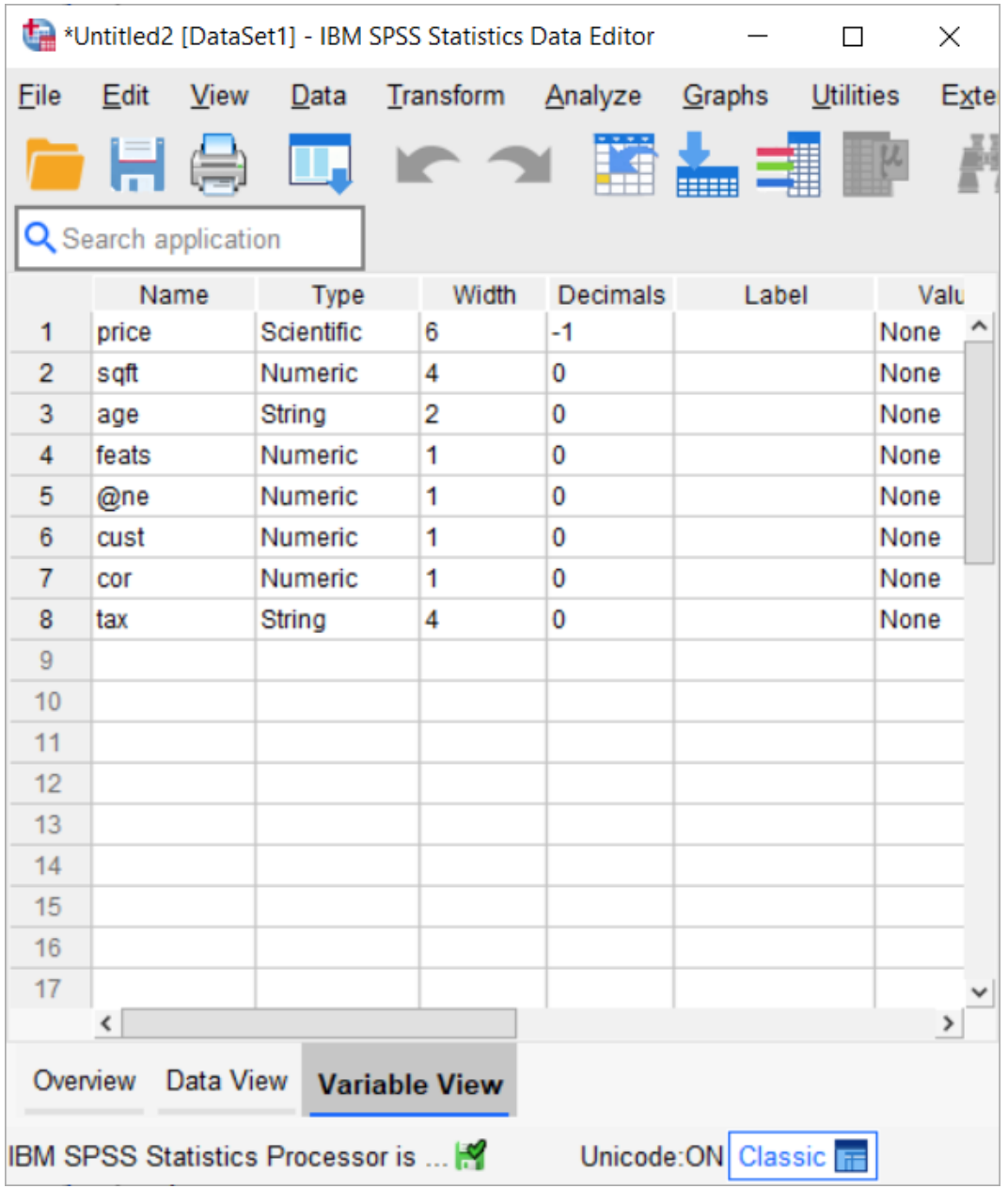

Figure 15: SPSS Variable View with default choices

Speaker notes

SPSS usually does pretty well with its default choices, but not here. There are several issues of immediate concern. SPSS used scientific notation for price, and failed to recognize age and tax as numbers, not strings. This was a side effect of coding missing values as NA. NA is a great choice for R, but SPSS (and SAS) both prefer a single dot. Finally, SPSS did not like the ne variable because it conflicts with the ne code for not equal to in SPSS syntax.

These are all easy to fix.

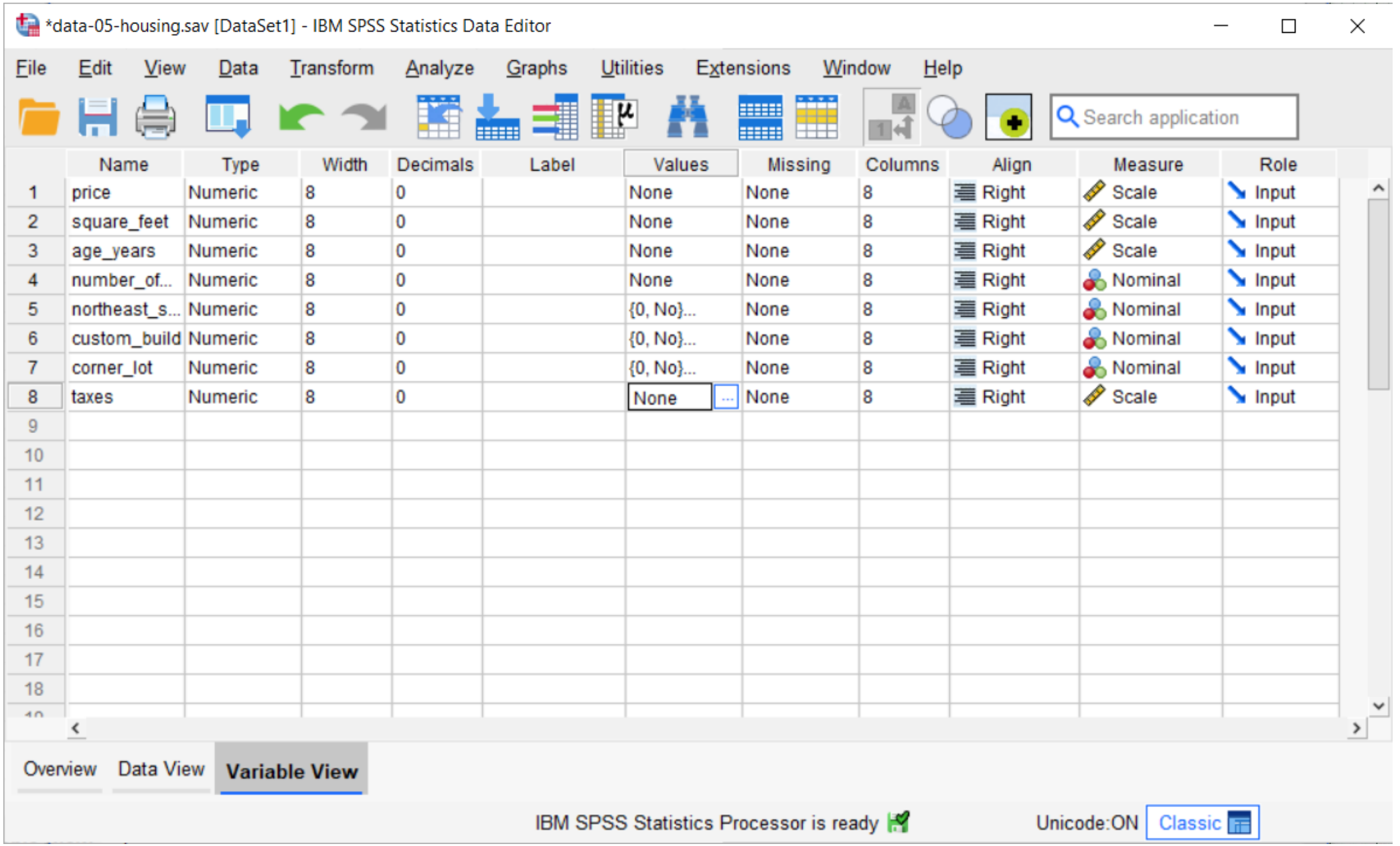

Figure 16: SPSS Variable View, updated

Speaker notes

This shows the updated view with the correct variable types and variable names. I also included value labels for the 0, 1 codes.

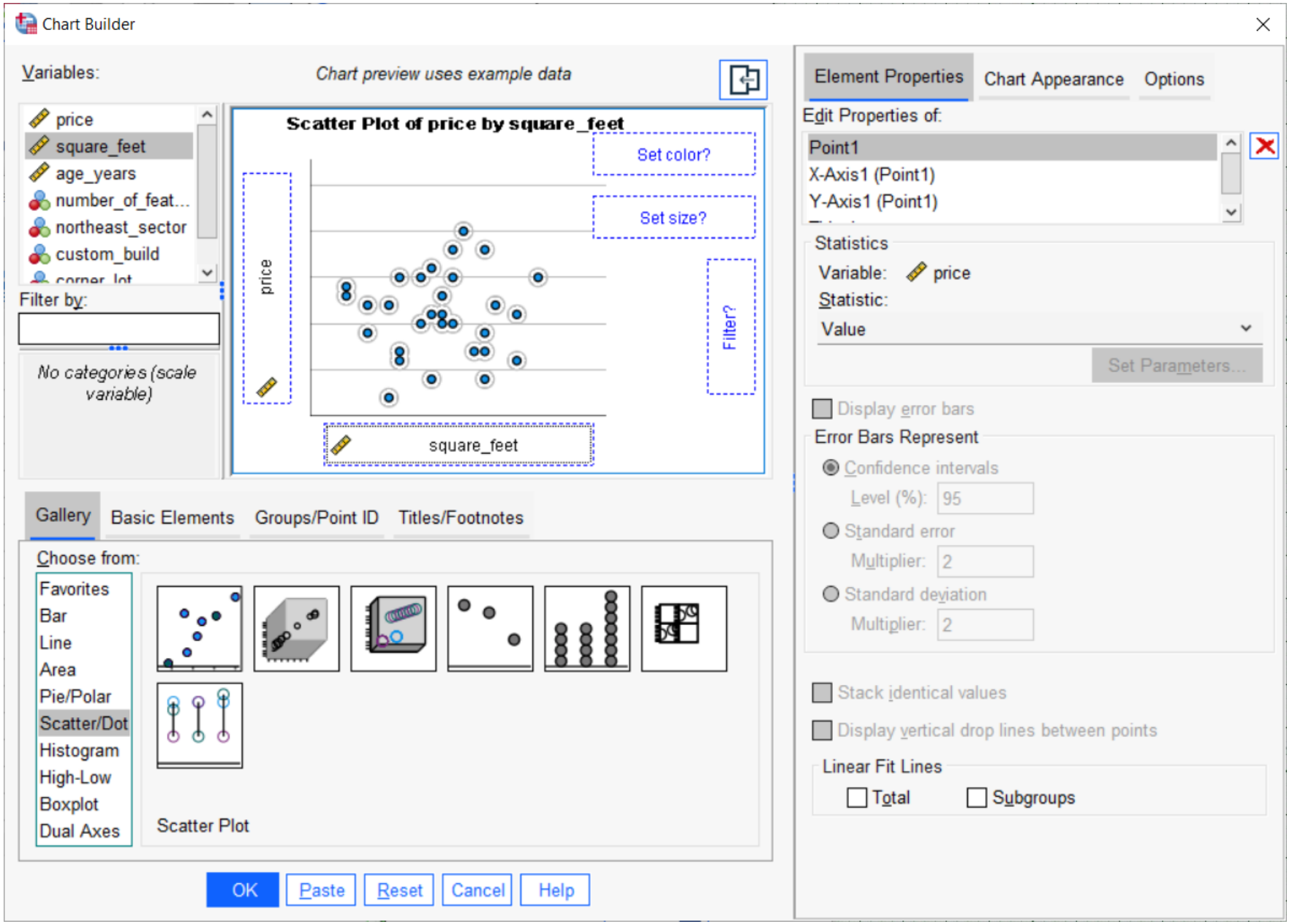

Figure 17: SPSS chart builder dialog box

Speaker notes

This is the chart builder dialog box for a simple scatterplot. Drag and drop the icon in the first row, first on the left.

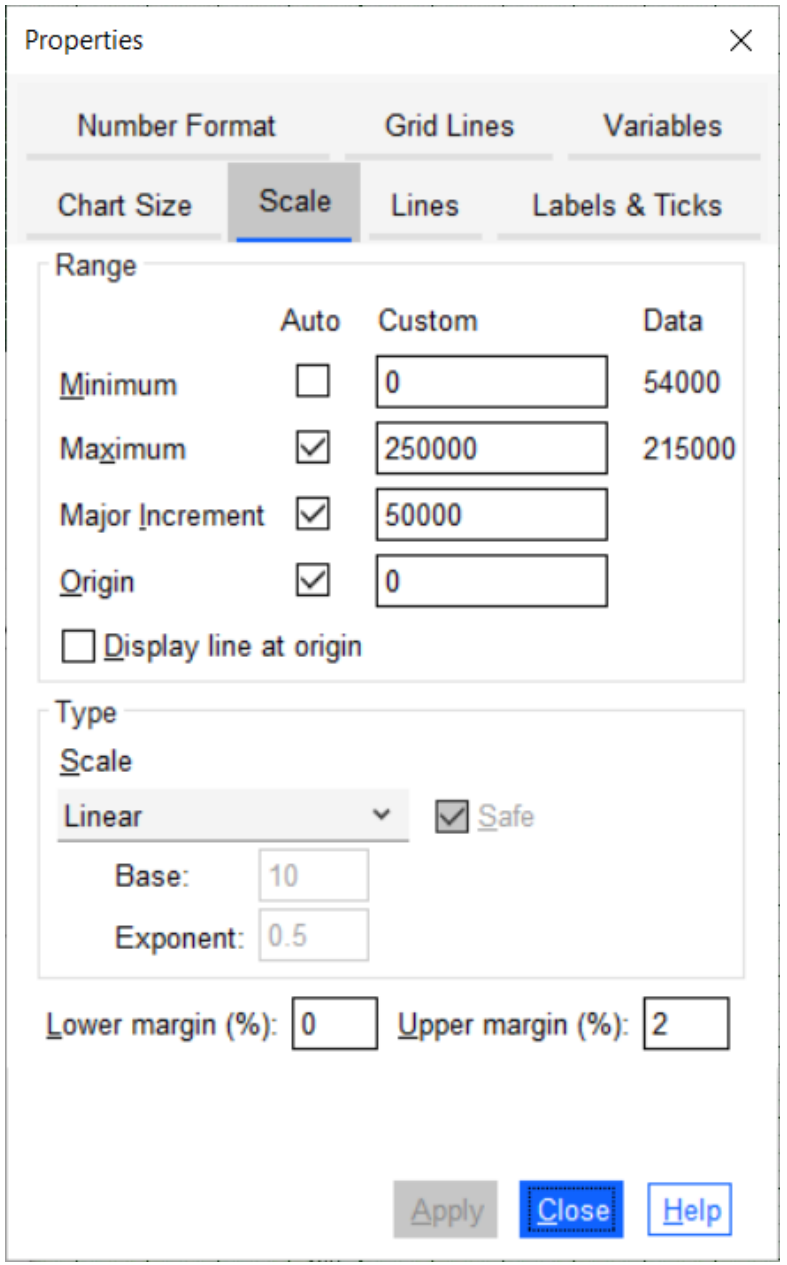

Figure 18: SPSS Properties dialog box, Scale tab

Speaker notes

Double click on the scatterplot to call up the properties dialog box. Then double click on the Y-axis. The scale tab allows you to include zero on the Y-axis.

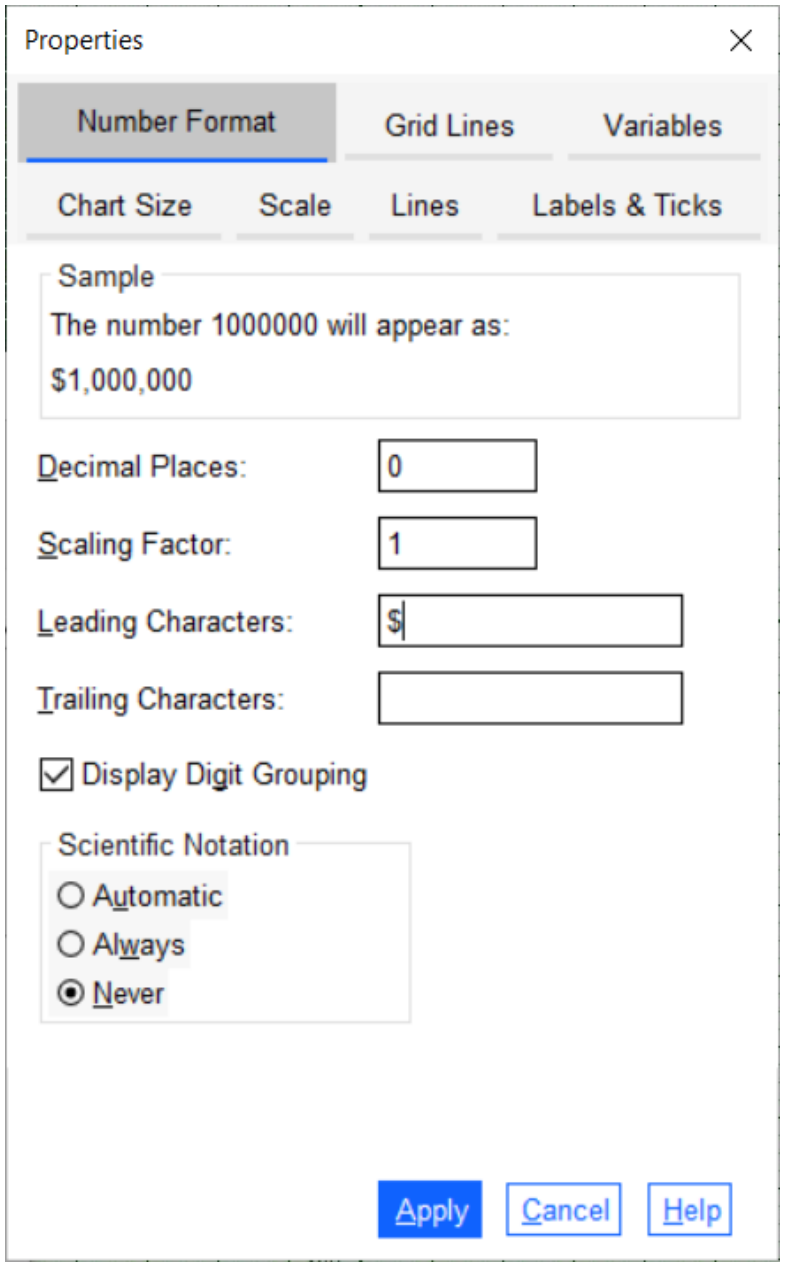

Figure 19: SPSS Properties dialog box, Number Format tab

Speaker notes

Click on the numbers format tab to make the numbers on this axis look nicer. This screenshot shows how to add a dollar sign to the front of the number and to use a comma separator.

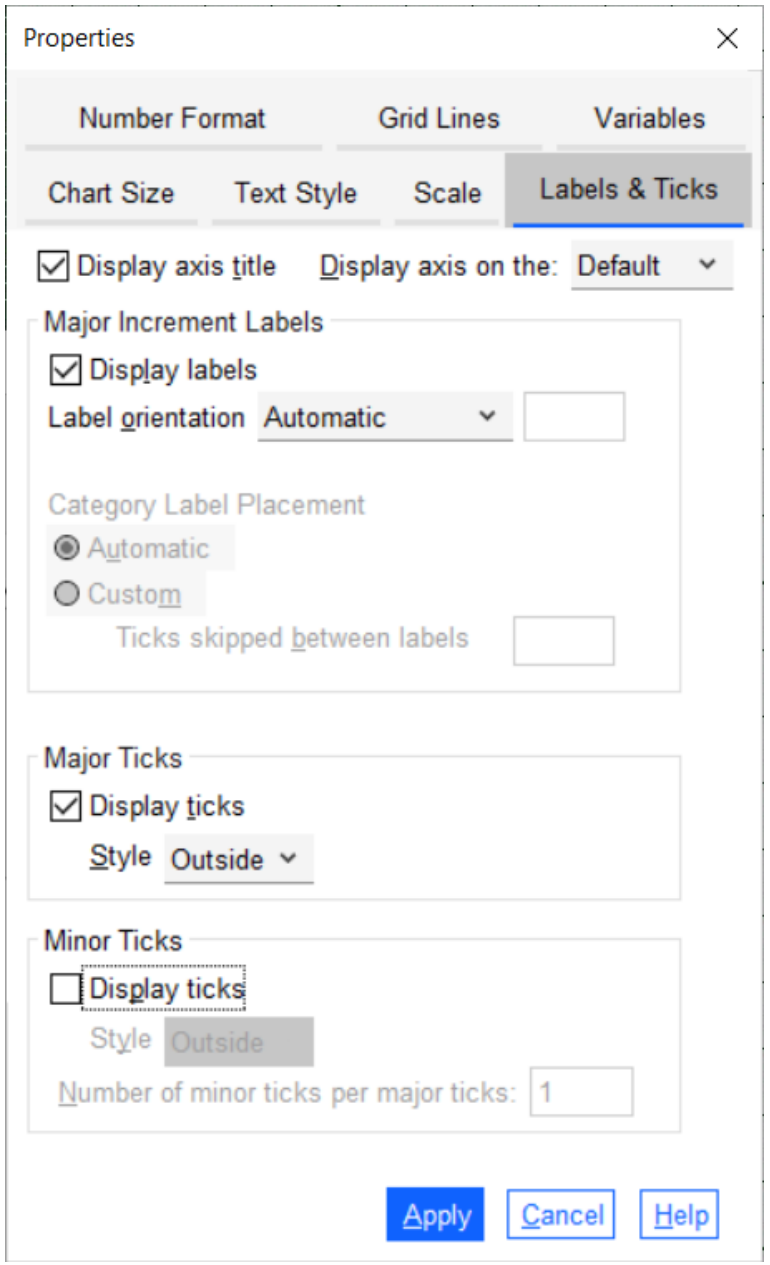

Figure 20: SPSS Properties dialog box. Labels & Ticks tab

Speaker notes

It’s a bit odd that SPSS does not use tick marks as a default option on scatterplots. You can update this in the Labels & Ticks tab.

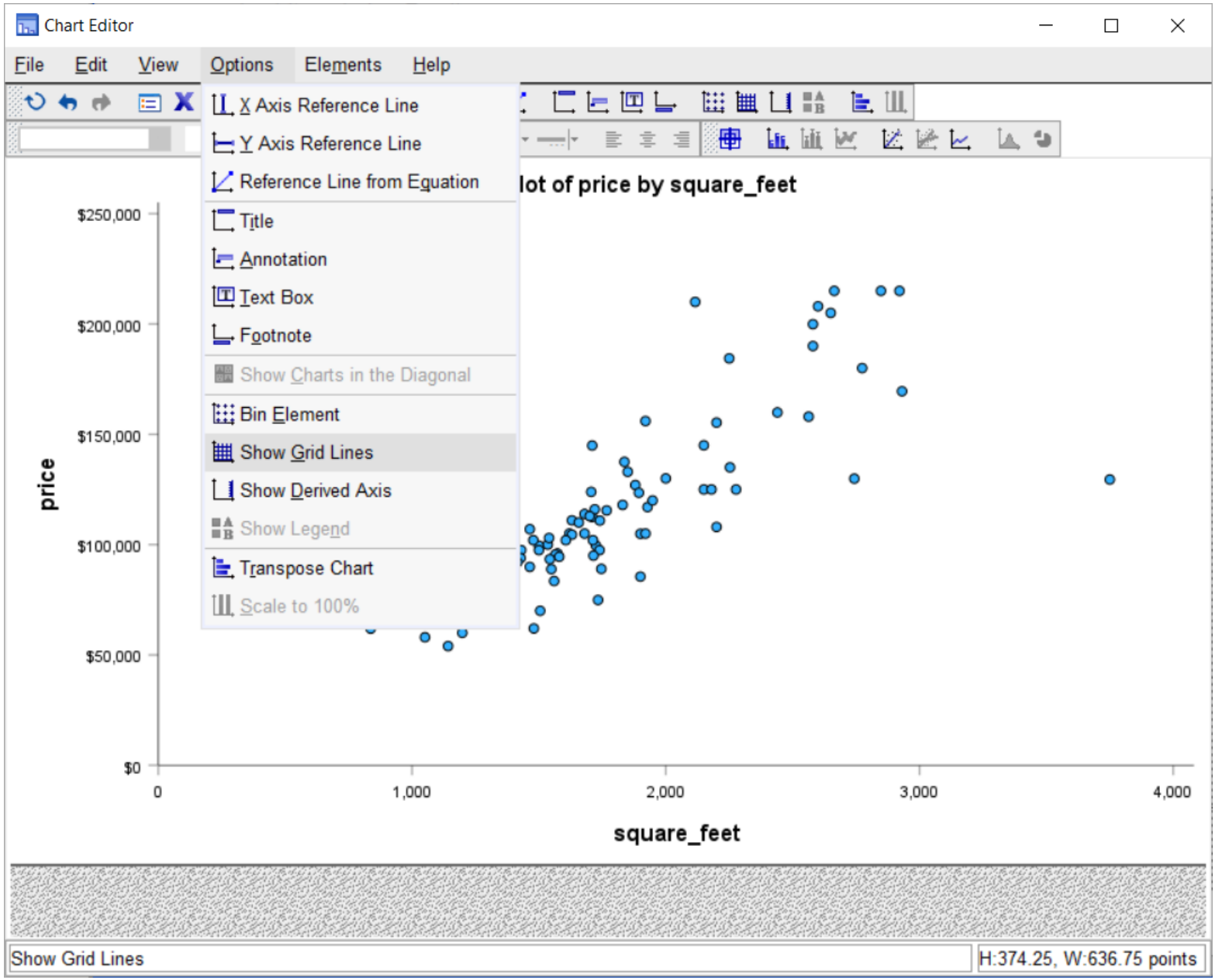

Figure 21: SPSS Chart Editor, Options menu

Speaker notes

Another puzzling default in SPSS is the use of grid lines for the Y-axis, but not the X-axis. You can modify this from the options menu.

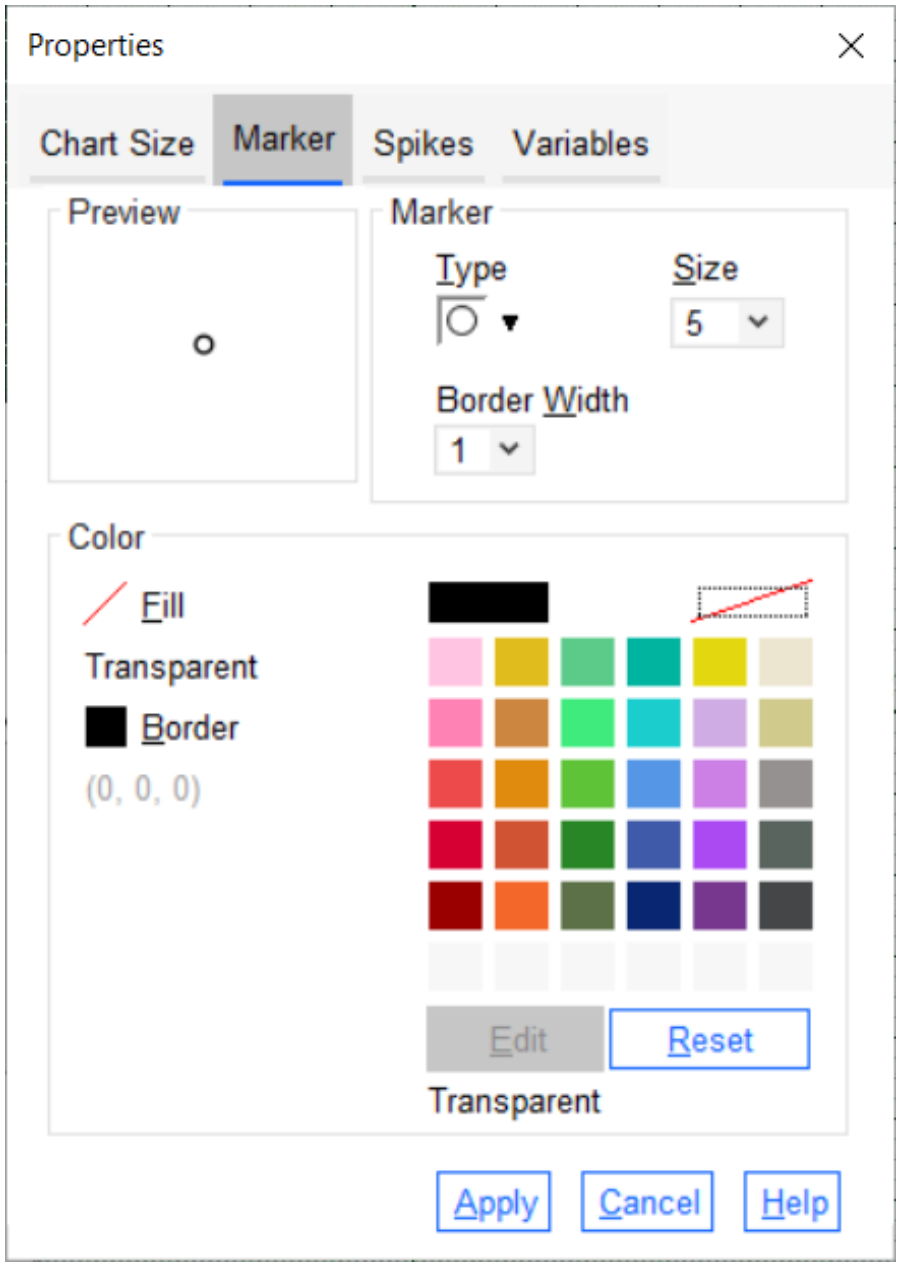

Figure 22: SPSS Properties dialog box

Speaker notes

Circles are good markers, but I prefer to see behind the color rather than having a filled in circle. You can change the fill color to transparent in the marker tab. A transparent fill helps a lot when there are a bunch of points almost on top of one another.

Break 3

- What have you learned?
  - Housing data, scatterplot
- What’s coming next
  - Housing data, boxplot

Speaker notes

Time for a break. Any questions?

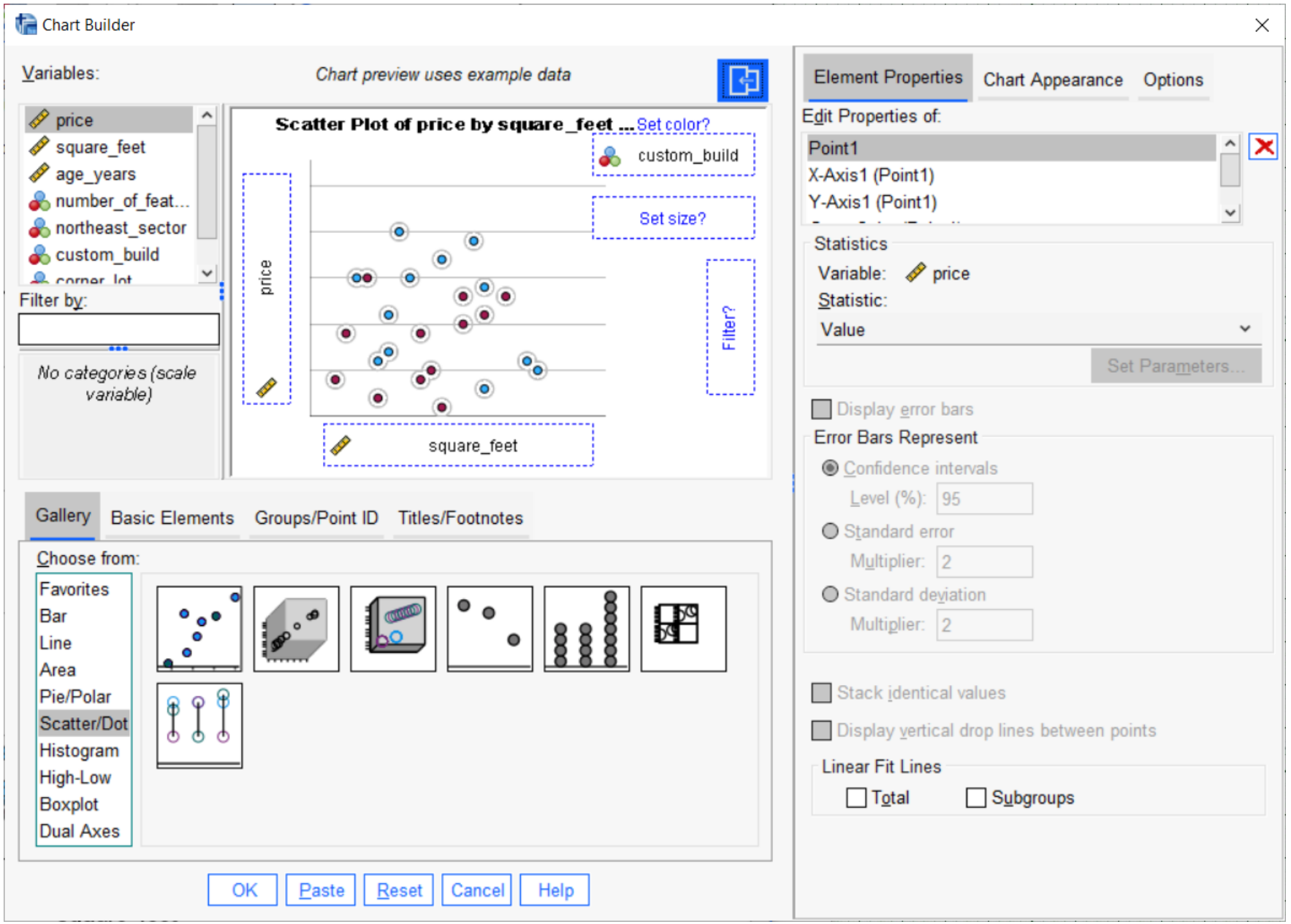

Figure 23: SPSS Chart Builder, scatterplot with colors

Speaker notes

You can add a third variable using color. Drag and drop the icon in the first row, fourth from the right.

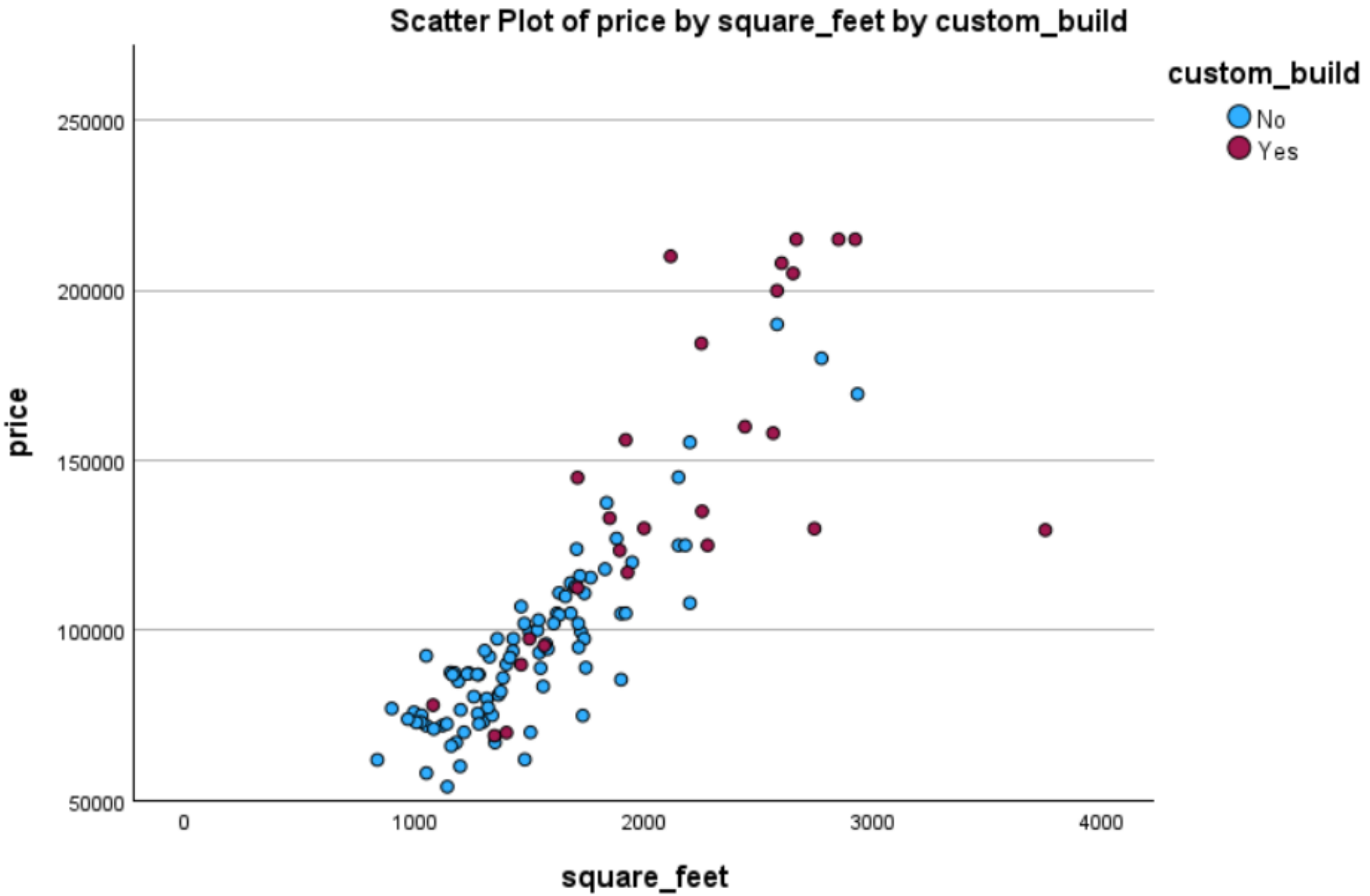

Figure 24: SPSS output

Speaker notes

This is what the plot looks like with the default options. It is not too bad, but could use some improvement. The trend between square footage and price is similar for the custom built and regular houses, though the custom built houses are a bit more expensive.

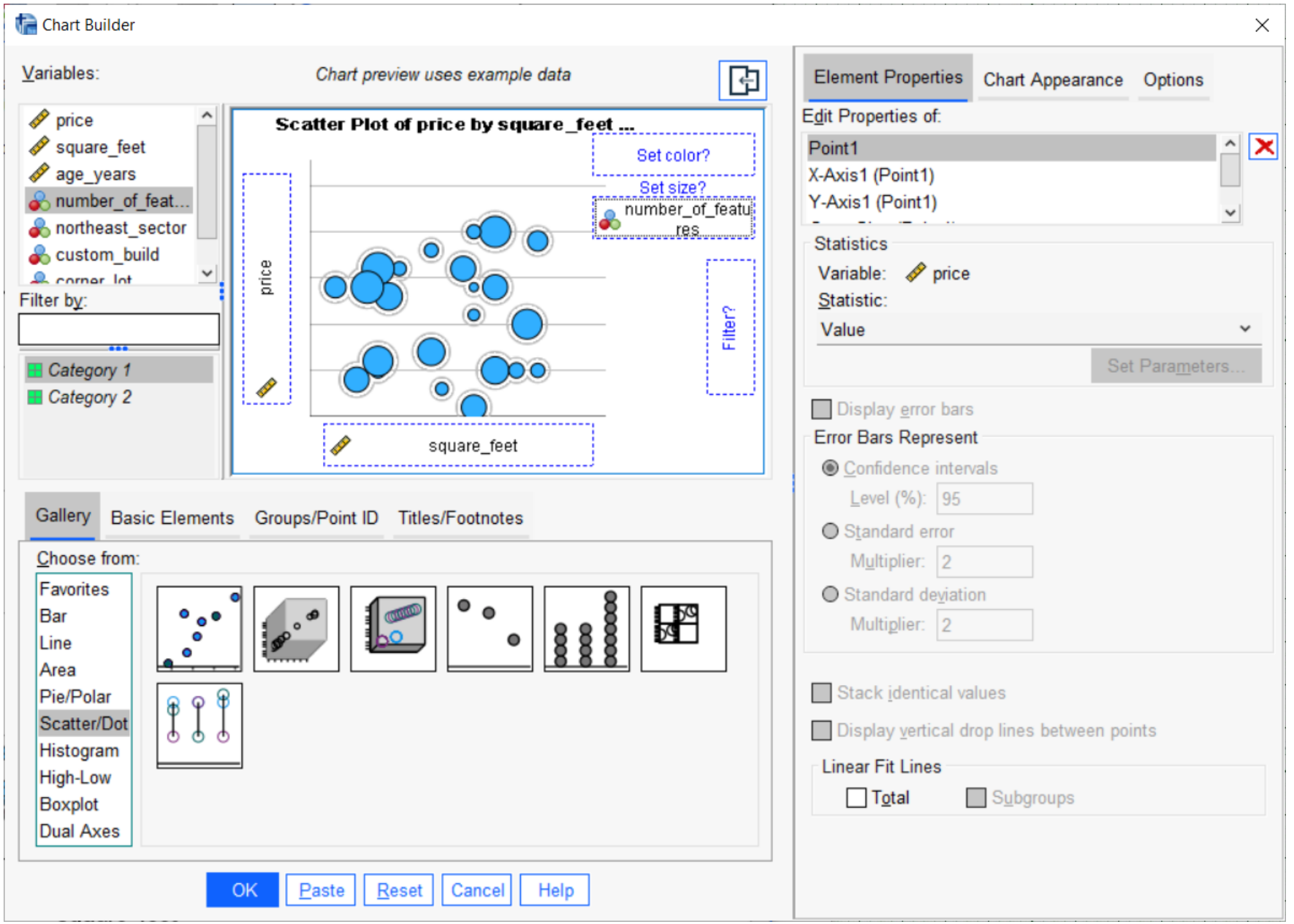

Figure 25: SPSS chart builder

Speaker notes

Add note.

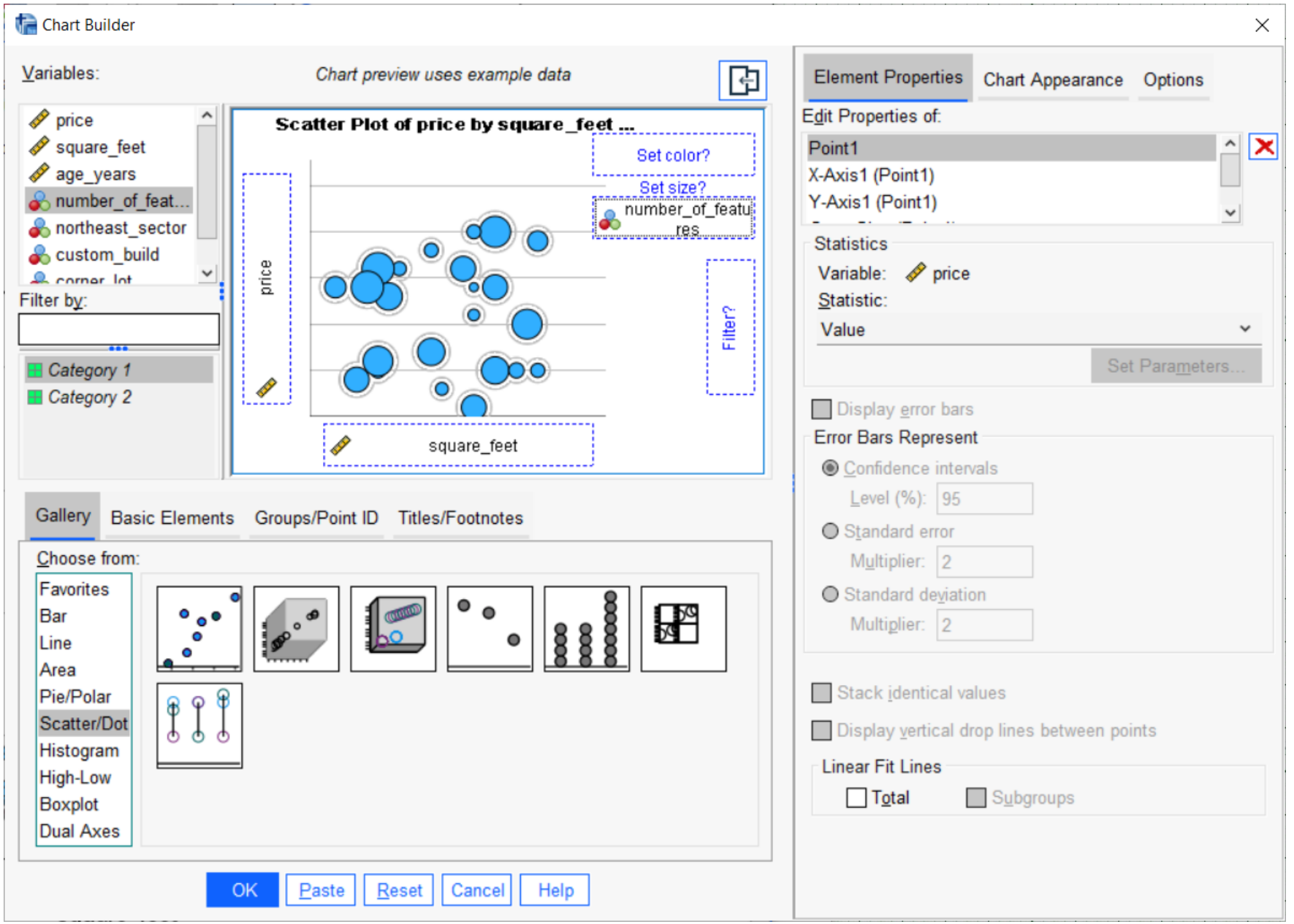

Figure 26: SPSS

Speaker notes

Add note.

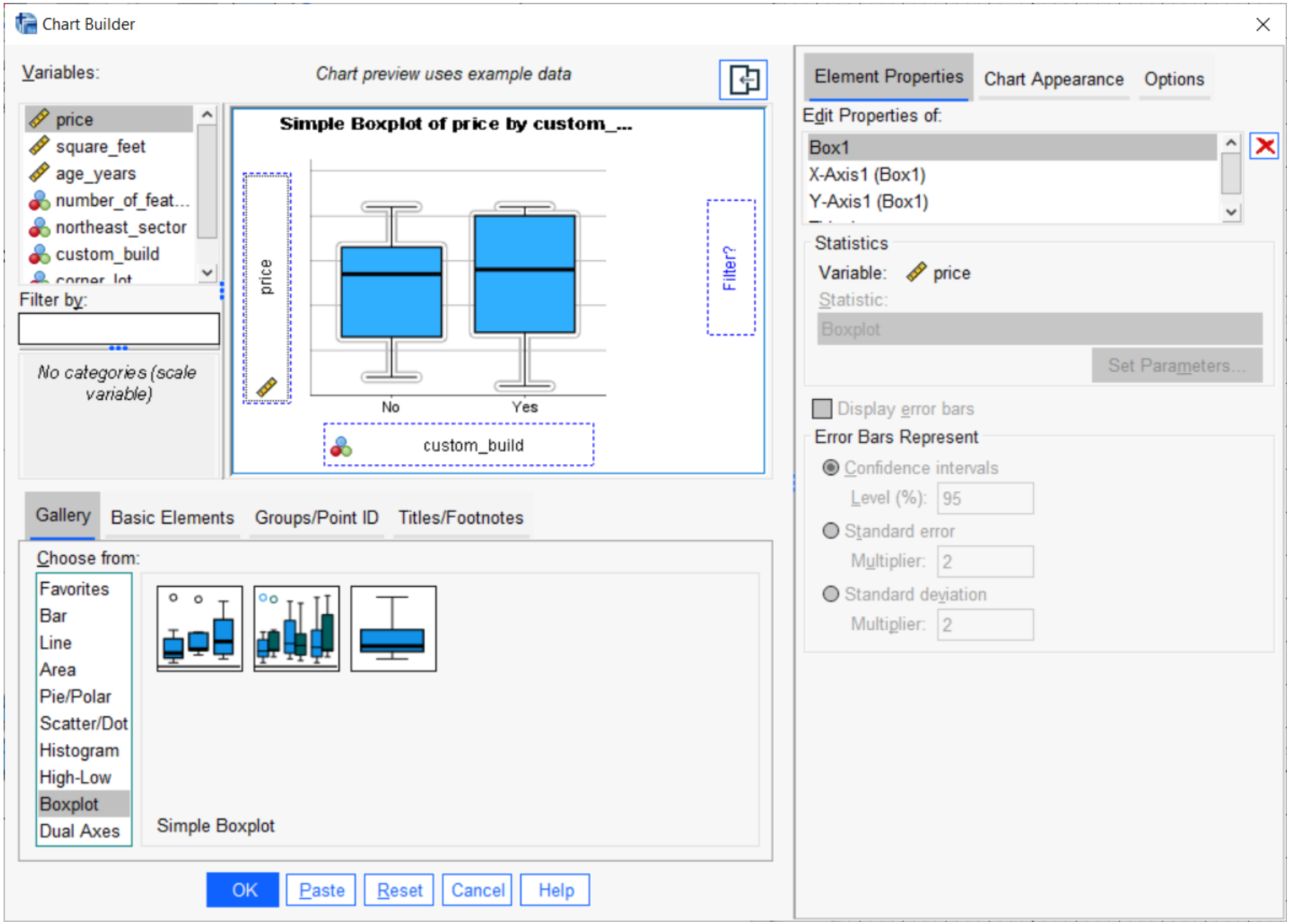

Figure 27: SPSS chart builder

Speaker notes

Add note.

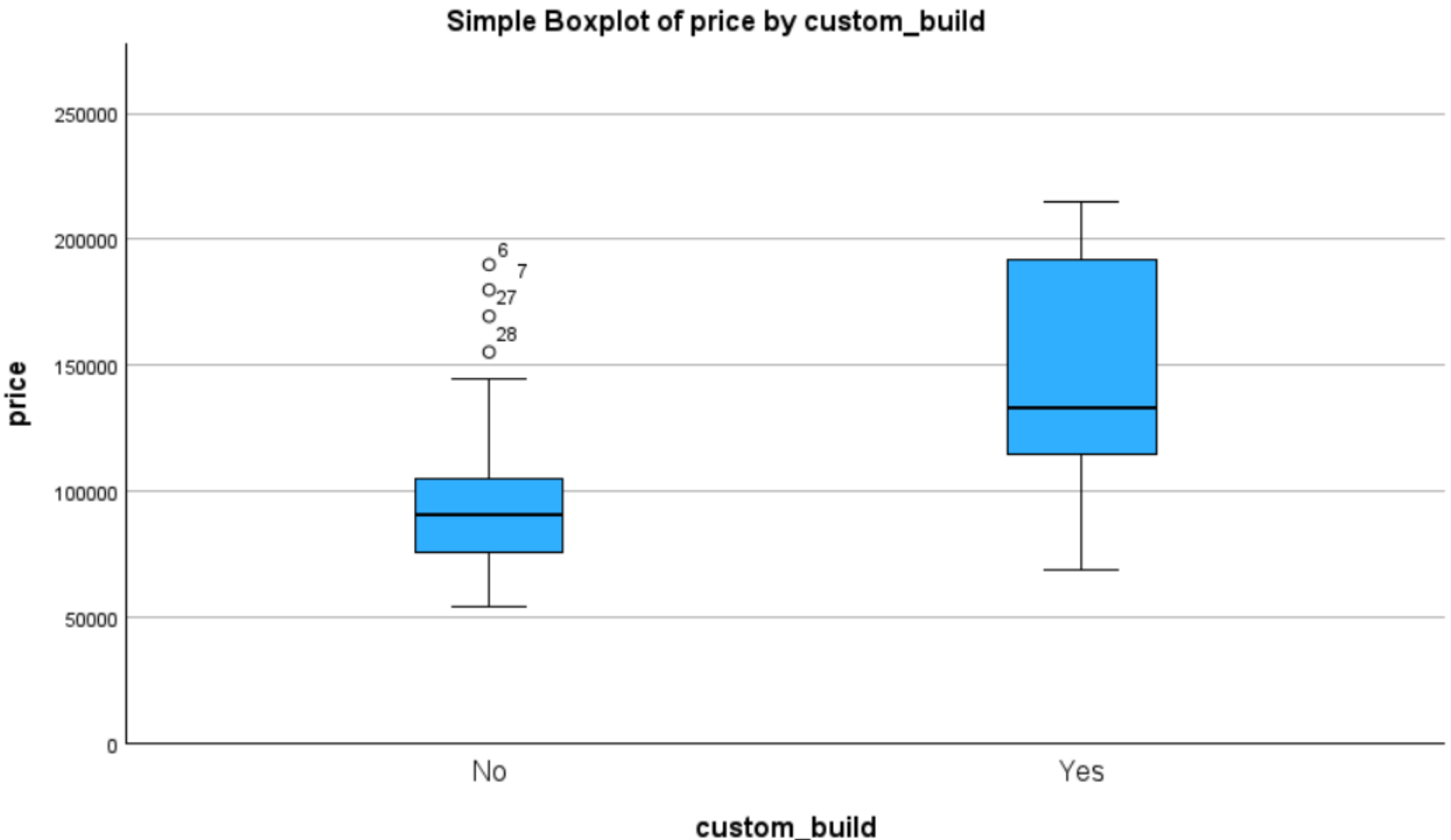

Figure 28: SPSS output

Speaker notes

Add note.

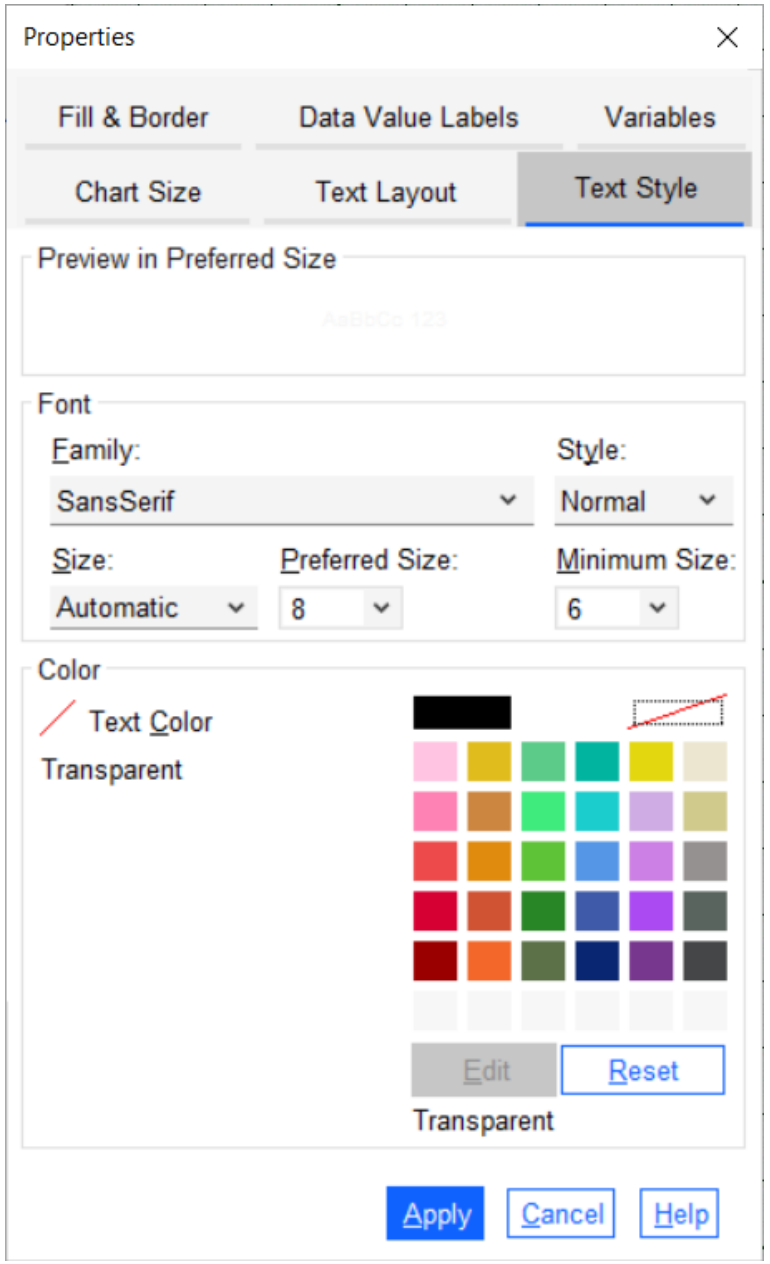

Figure 29: SPSS Properties dialog box

Speaker notes

Add note.

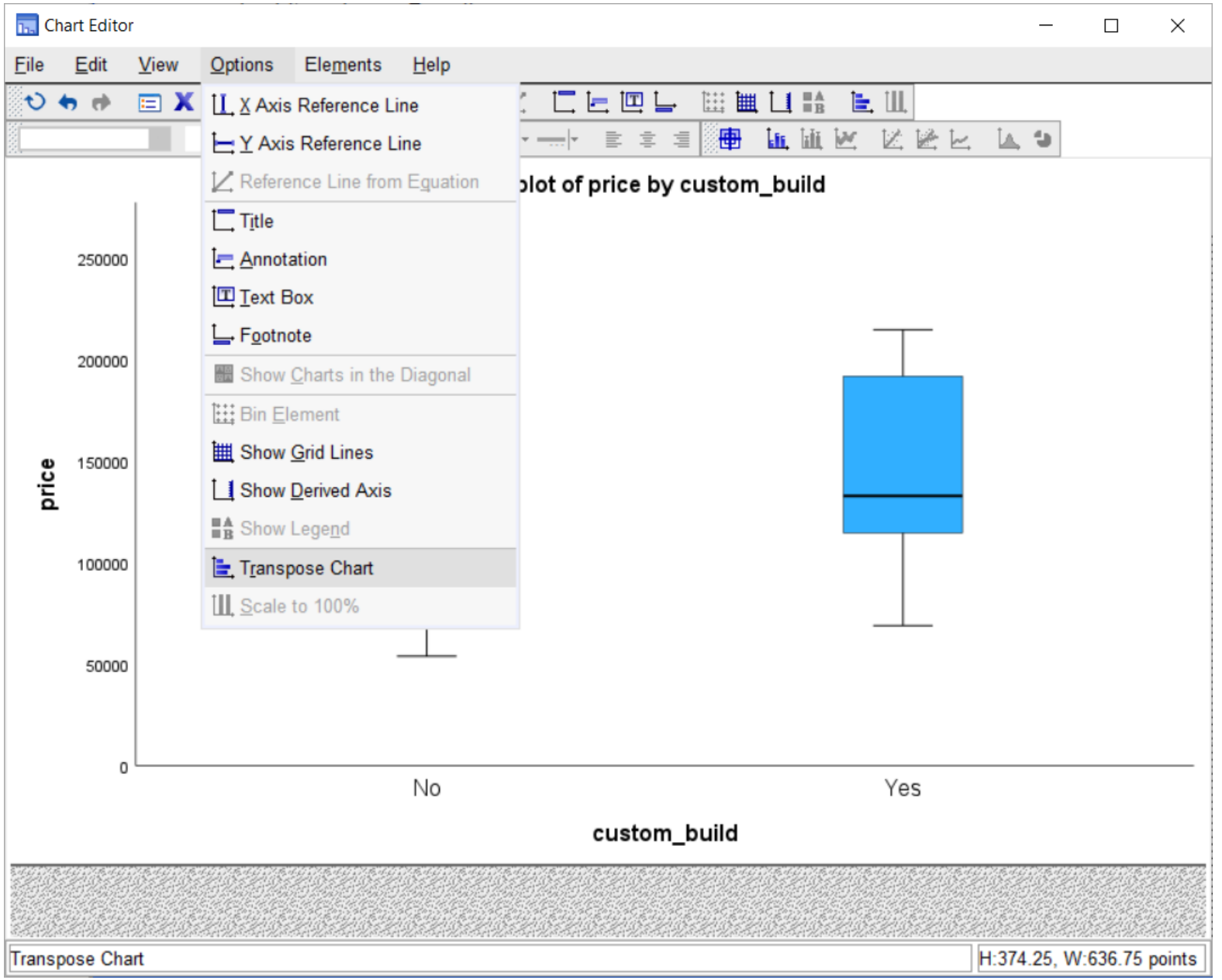

Figure 30: SPSS options menu

Speaker notes

Add note.

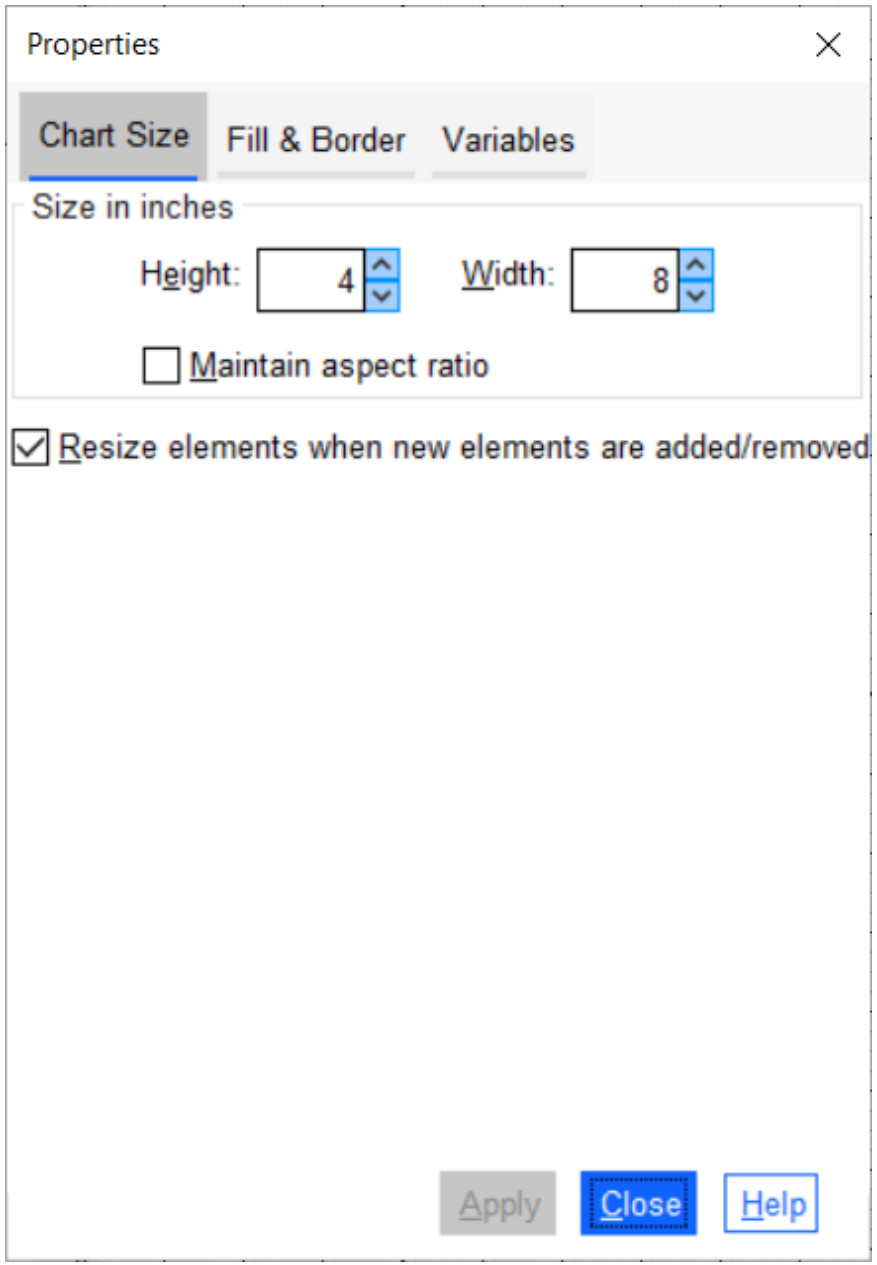

Figure 31: SPSS Properties dialog box

Speaker notes

Add note.

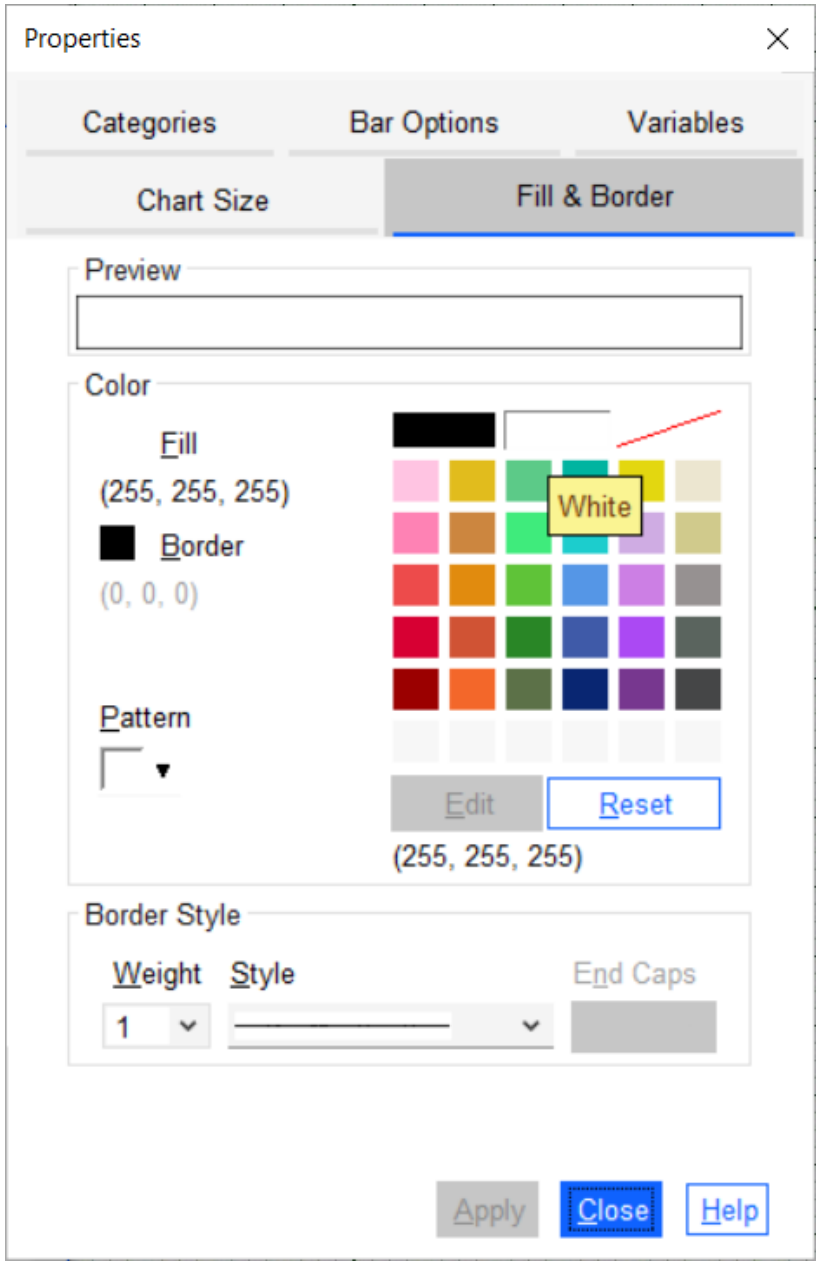

Figure 32: SPSS Properties dialog box

Speaker notes

Add note.

Figure 33: SPSS output

Speaker notes

Add note.

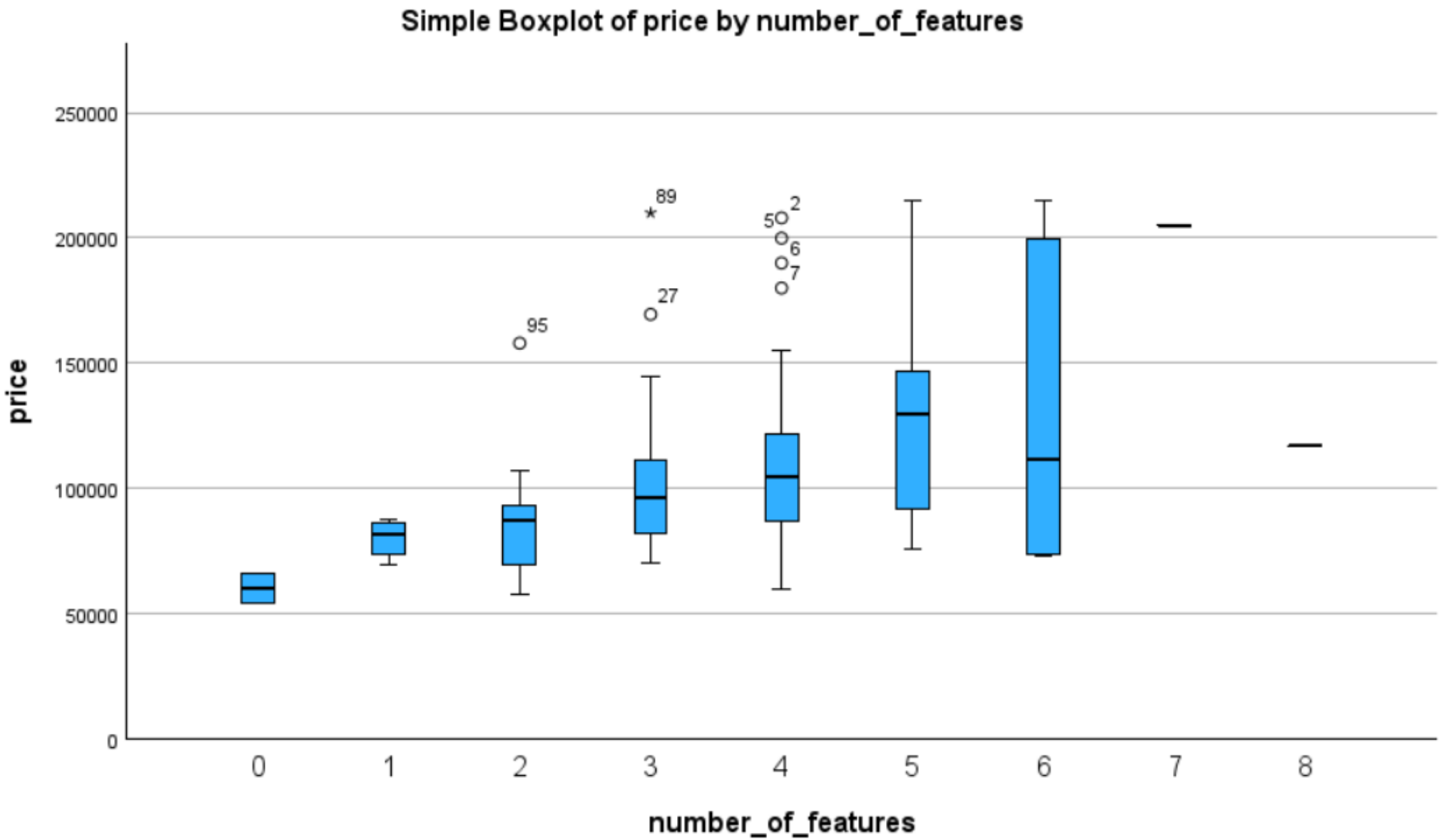

Figure 34: SPSS output

Speaker notes

Add note.

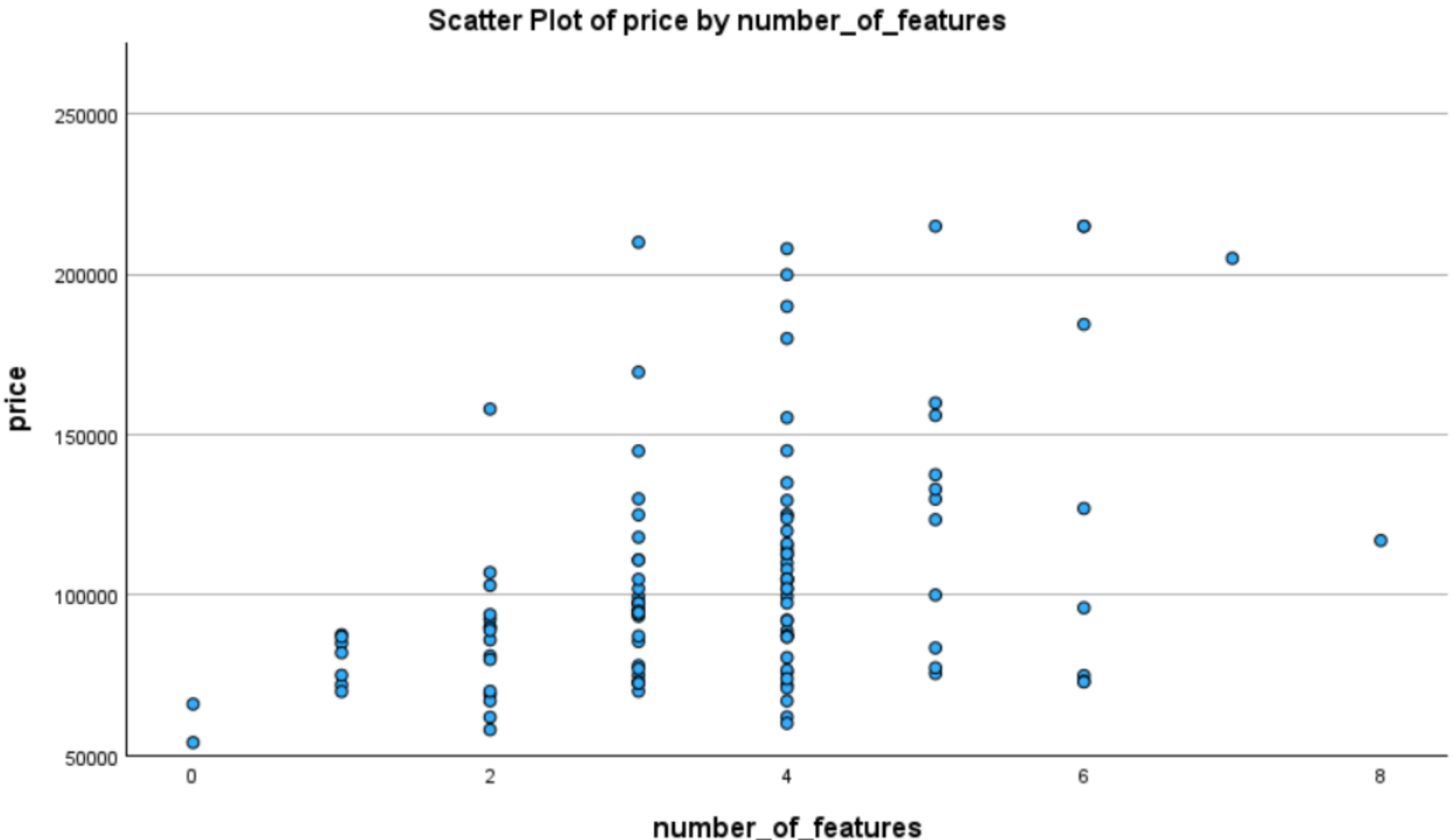

Figure 35: SPSS output

Speaker notes

Add note.

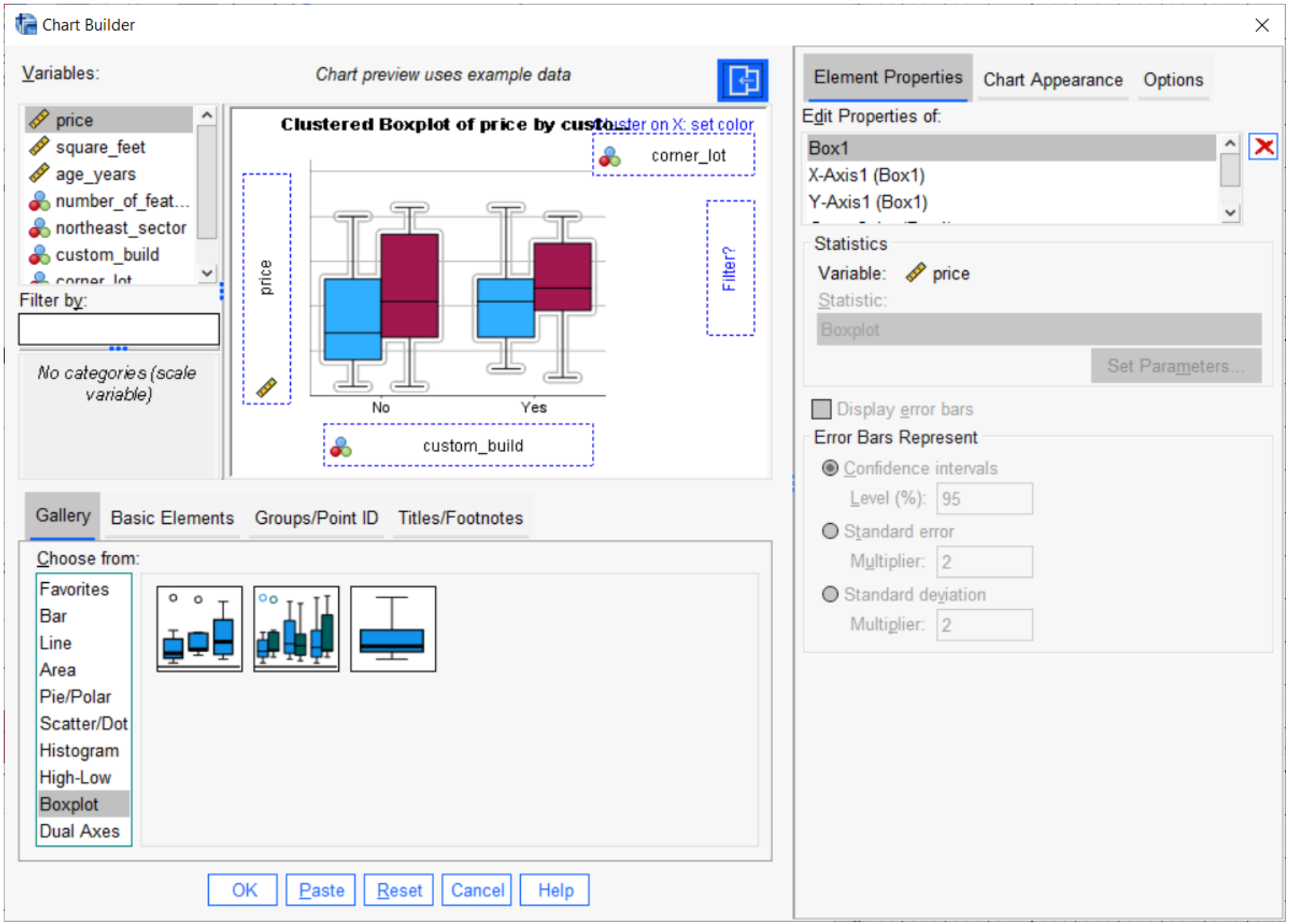

Figure 36: SPSS chart builder

Speaker notes

Add note.

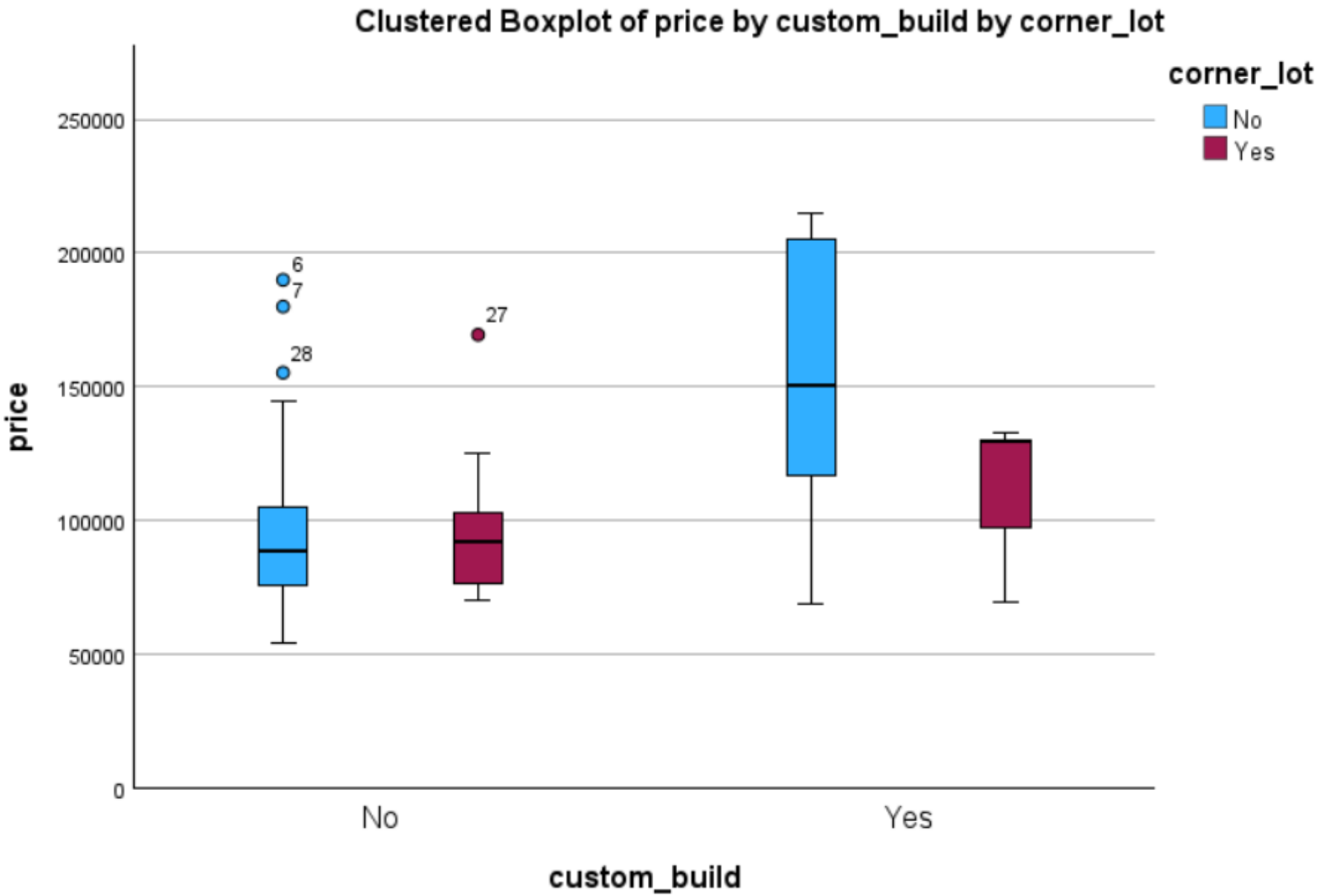

Figure 37: SPSS output

Speaker notes

Add note.