

How to use the Turnip Graph Engine in STATA

WHAT YOU HAVE TO DO:

1. Download `turnip.ado` and put it in an ado file folder
2. Now whenever you open STATA you can run `turnip` as you would any `.ado` file: for example, if you have a data file open with a variable called `mpg`, just type
`turnip mpg`
3. If you prefer, you can change the name of the file to `turnip.do` and run it as you would any `.do` file (i.e., open STATA, run `turnip.do`, then you can use `turnip` by typing `turnip` and the name of your variable).

SYNTAX

```
turnip varlist [if exp] [, RESolution() TRUev() GraphicsOptions]
```

Description

Creates a turnip-style graph. Turnip graphs display distributions like a boxplot; but, because they display the actual data points, they are much easier to understand. Plus, they look really cool.

Options

All standard graphics options except for "BY" are allowed. If you want to align multiple turnips you can use the STATA function: `^gr combine^` (see `@gr combine.hlp@`).

Options specific to TURNIP:

1. `resolution` - Default resolution of the graphic (i.e., how many categories `varlist` is divided into) is $.4 * (\text{standard deviation of } \text{'varlist'})$ variable. Alternatively, user can specify resolution. Since the resolution feature rounds the data, the graphic in essence displays the frequency of observations falling within in resolution unit. The user can avoid any such rounding (i.e., display frequency of each value in the data) by specifying a negative number in the resolution option. For example, `turnip xxx, res(-1)`
2. `xlabel` - Default `xlabel` is twice the maximum number of observations in any `tt`varlist'category`. User can reset `xlabel` to adjust width of the graphic.
3. `ylabel` - Default `ylabel` adds 25% of the range above and below min and max values. User can reset `ylabel` to adjust height of graphic.

4. truev - The program rounds values of the varlist variable into units defined by the resolution option. In some cases the user may want to prevent this from happening. For example, suppose you are trying to display change scores and want to show which observations are above or below zero . The truev (as in true value) option ensures that only zero values are graphed at zero; other values which would round to zero are set at the appropriate next most category of the varlist variable.
5. NOTE: In addition to the usual function of YLINE, turnip allows user to specify special values; specifically, mean or median. If either mean or median is specified, the corresponding value is added to ylabel so the value is displayed on the y-axis.

ACKNOWLEDGEMENT

Turnips are the graphic of choice in the Dartmouth Atlas of Healthcare (American Hospital Publishing Inc., Chicago, IL, 1996). The term "turnip" graph was coined by Jack Wennberg, MD, MPH and colleagues at the Center for Evaluative Clinical Sciences, Dartmouth Medical School.

EXAMPLES

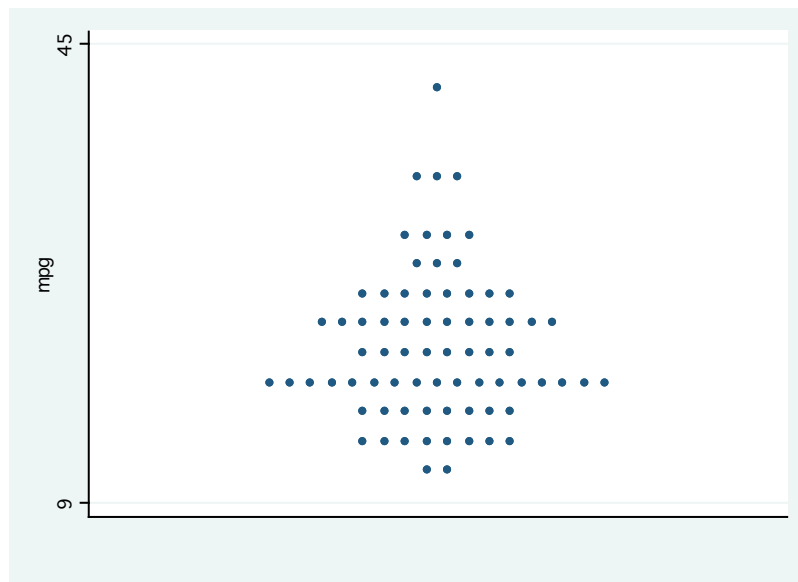
I. Plain

```
use ":\Macintosh HD:Applications:Stata 6.0:auto.dta"  
(1978 Automobile Data)
```

```
. turnip mpg  
(74 real changes made)  
(39 real changes made)  
# of observations per turnip row is the frequency below:
```

mpg	Freq.	Percent	Cum.
11.55	2	2.70	2.70
13.86	8	10.81	13.51
16.17	8	10.81	24.32
18.48	17	22.97	47.30
20.79	8	10.81	58.11
23.1	12	16.22	74.32
25.41	8	10.81	85.14
27.72	3	4.05	89.19
30.03	4	5.41	94.59
34.65	3	4.05	98.65
41.58	1	1.35	100.00
Total	74	100.00	

```
options: xlabel(-16 16) ylabel(9 45 )  
resolution: 2.31  
truev: 1.00e+11  
dflag = none
```



II. Suspend rounding

```
turnip mpg, res(-1)
```

*specifying negative number for resolution turns off all rounding.

```
. turnip mpg, res(-1)
```

```
(74 real changes made)
```

```
(43 real changes made)
```

of observations per turnip row is the frequency below:

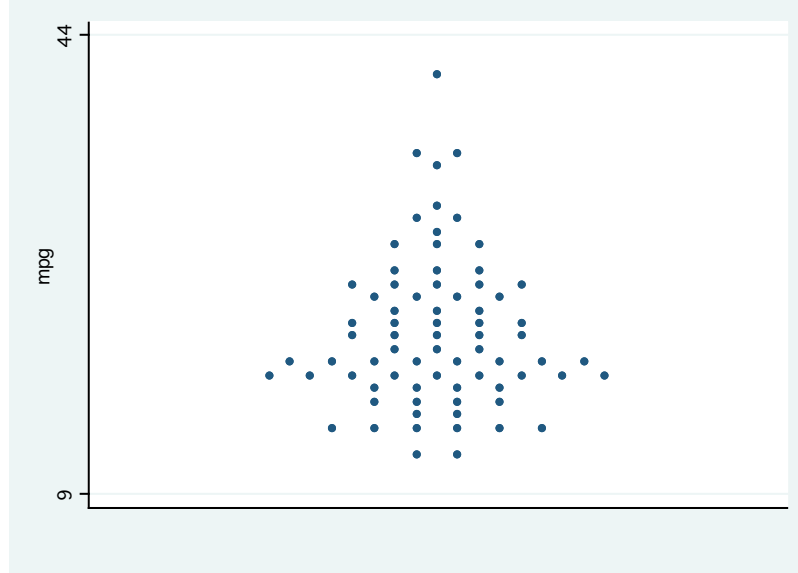
mpg	Freq.	Percent	Cum.
12	2	2.70	2.70
14	6	8.11	10.81
15	2	2.70	13.51
16	4	5.41	18.92
17	4	5.41	24.32
18	9	12.16	36.49
19	8	10.81	47.30
20	3	4.05	51.35
21	5	6.76	58.11
22	5	6.76	64.86
23	3	4.05	68.92
24	4	5.41	74.32
25	5	6.76	81.08
26	3	4.05	85.14
28	3	4.05	89.19
29	1	1.35	90.54
30	2	2.70	93.24
31	1	1.35	94.59
34	1	1.35	95.95
35	2	2.70	98.65
41	1	1.35	100.00
Total	74	100.00	

```
options: xlabel(-8 8) ylabel(9 44)
```

```
resolution: -1
```

```
truev: 1.00e+11
```

```
dflag = none
```



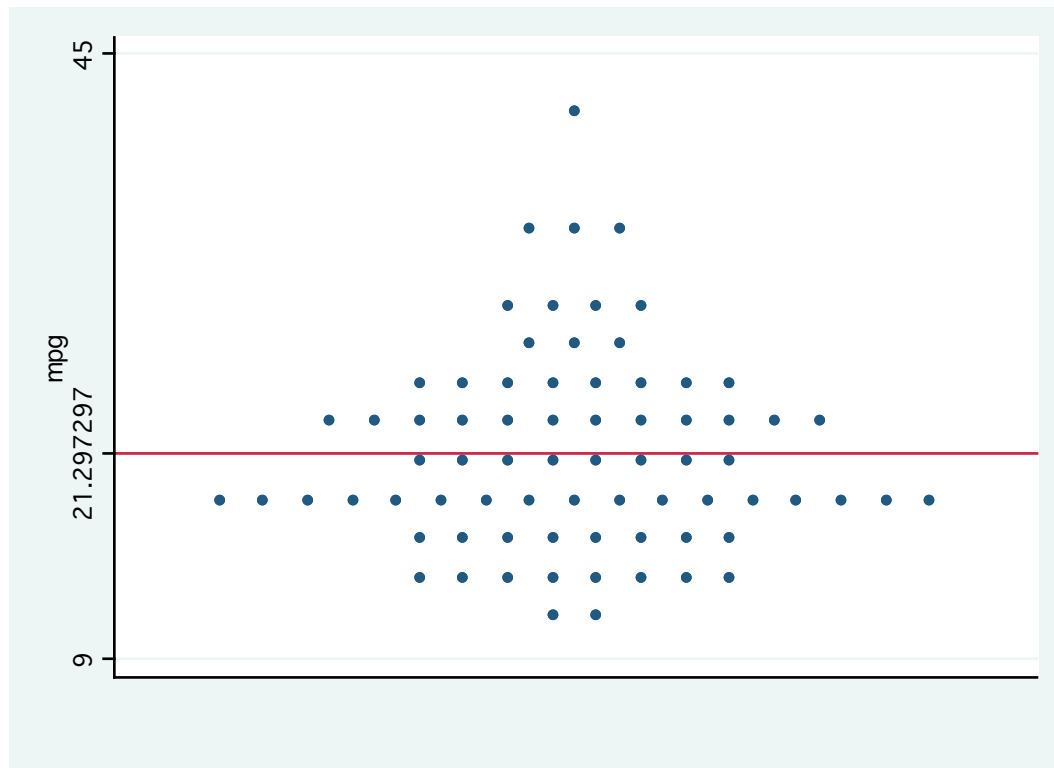
III. Specify xlabel (controls turnip width) and add a yline

```
turnip mpg, xlabel(-10 10) yline(mean)
*ylines(mean) adds a line at the distribution's mean and automatically
adds the value for the mean to the y-label
```

```
. turnip mpg, xlabel(-10 10) yline(mean)
(74 real changes made)
(39 real changes made)
# of observations per turnip row is the frequency below:
```

mpg	Freq.	Percent	Cum.
11.55	2	2.70	2.70
13.86	8	10.81	13.51
16.17	8	10.81	24.32
18.48	17	22.97	47.30
20.79	8	10.81	58.11
23.1	12	16.22	74.32
25.41	8	10.81	85.14
27.72	3	4.05	89.19
30.03	4	5.41	94.59
34.65	3	4.05	98.65
41.58	1	1.35	100.00
Total	74	100.00	

```
options: xlabel(-10 10) yline(21.2) ylabel( 21.2972972972973 9 45 )
resolution: 2.31
truev: 1.00e+11
dflag = mean
ylines: 21.2972972972973 (mean)
```



IV. Use xlabel to narrow the graph (compare width to graph III)

```
turnip mpg, xlabel(-30 30) yline(mean)
```

```
. turnip mpg, xlabel(-30 30) yline(mean)
(74 real changes made)
(39 real changes made)
# of observations per turnip row is the frequency below:
```

mpg	Freq.	Percent	Cum.
11.55	2	2.70	2.70
13.86	8	10.81	13.51
16.17	8	10.81	24.32
18.48	17	22.97	47.30
20.79	8	10.81	58.11
23.1	12	16.22	74.32
25.41	8	10.81	85.14
27.72	3	4.05	89.19
30.03	4	5.41	94.59
34.65	3	4.05	98.65
41.58	1	1.35	100.00
Total	74	100.00	

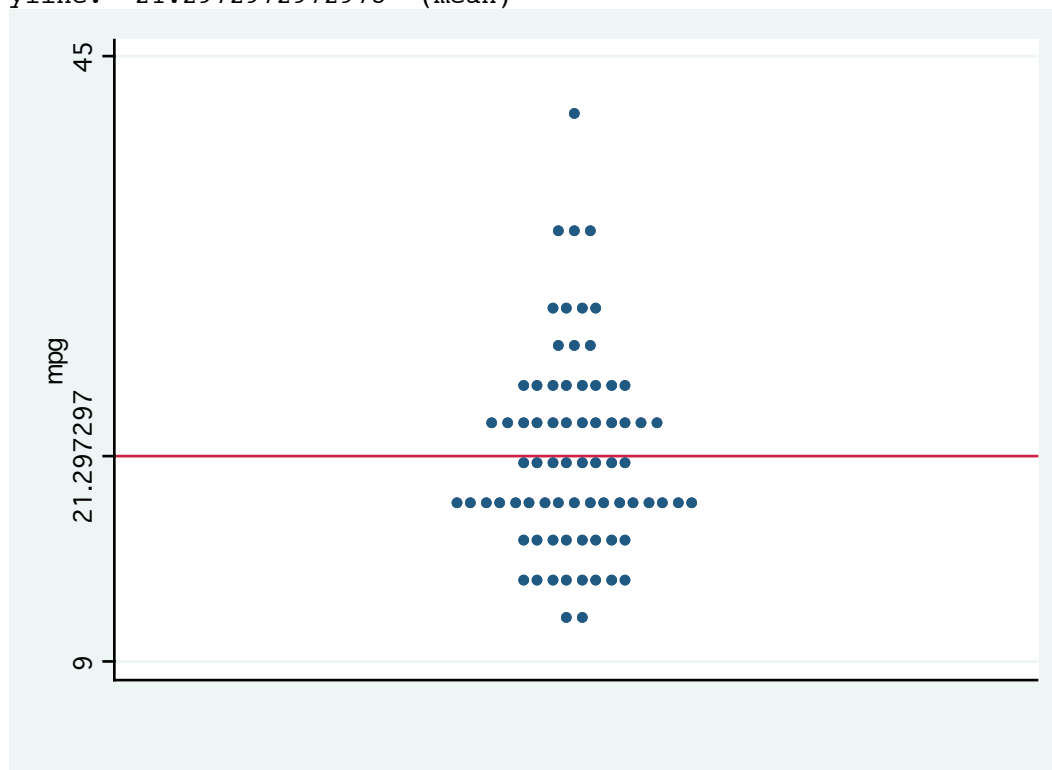
```
options: xlabel(-30 30) yline(21.2) ylabel( 21.2972972972973 9 45 )
```

```
resolution: 2.31
```

```
truev: 1.00e+11
```

```
dflag = mean
```

```
yline: 21.2972972972973 (mean)
```



V. Use "if" command

```
turnip mpg if foreign==1, xlabel(-10 10) yline(mean)
*IF command restricts graphic to only a those observations with
foreign==1
```

```
turnip mpg if foreign==1, xlabel(-10 10) yline(mean)
```

(22 real changes made)

(13 real changes made)

of observations per turnip row is the frequency below:

mpg	Freq.	Percent	Cum.
13.86	1	4.55	4.55
16.17	2	9.09	13.64
18.48	2	9.09	22.73
20.79	2	9.09	31.82
23.1	4	18.18	50.00
25.41	5	22.73	72.73
27.72	1	4.55	77.27
30.03	2	9.09	86.36
34.65	2	9.09	95.45
41.58	1	4.55	100.00
Total	22	100.00	

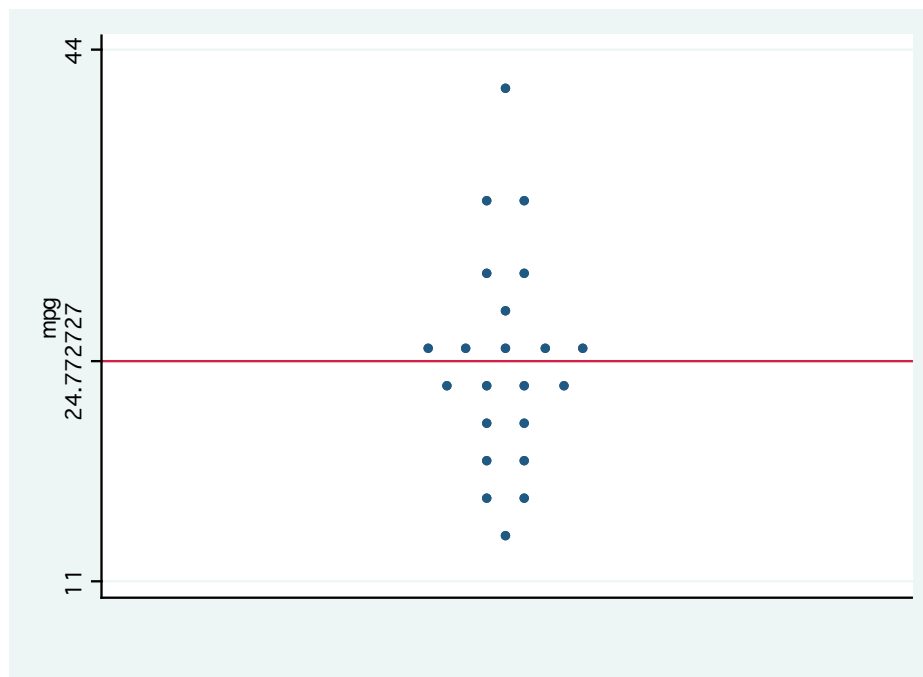
```
options: xlabel(-10 10) yline(24.7) ylabel( 24.77272727272727 11 44 )
```

```
resolution: 2.31
```

```
trueev: 1.00e+11
```

```
dflag = mean
```

```
ylines: 24.77272727272727 (mean)
```



VI. Use ylabel to narrow the graph to control height of graph (compare height to graph III)

```
turnip mpg , xlabel(-10 10) yline(mean) ylabel(10 100)
```

(74 real changes made)

(39 real changes made)

of observations per turnip row is the frequency below:

mpg	Freq.	Percent	Cum.
11.55	2	2.70	2.70
13.86	8	10.81	13.51
16.17	8	10.81	24.32
18.48	17	22.97	47.30
20.79	8	10.81	58.11
23.1	12	16.22	74.32
25.41	8	10.81	85.14
27.72	3	4.05	89.19
30.03	4	5.41	94.59
34.65	3	4.05	98.65
41.58	1	1.35	100.00
Total	74	100.00	

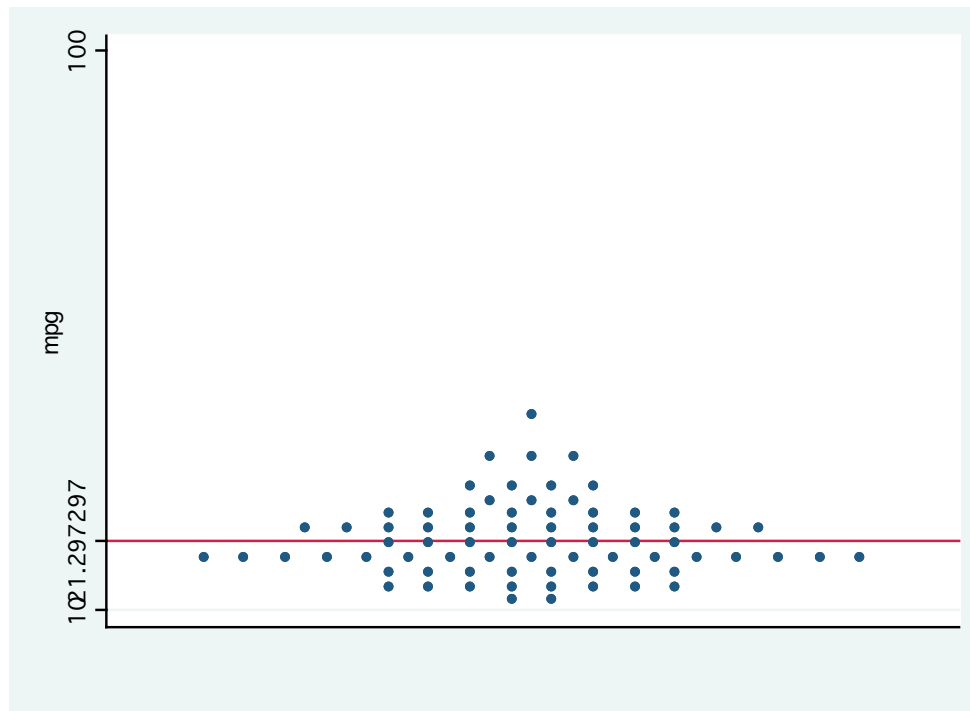
```
options: xlabel(-10 10) yline(21.2) ylabel( 21.2972972972973 10 100)
```

```
resolution: 2.31
```

```
truev: 1.00e+11
```

```
dflag = mean
```

```
ylines: 21.2972972972973 (mean)
```



VII. Round y values

```
turnip mpg , xlabel(-10 10) yline(mean) ylabel(10 50) res(2)  
*rounds y'values to the nearest 2 (eg, in table below with table in ex.  
VI).
```

```
(74 real changes made)
```

```
(41 real changes made)
```

```
# of observations per turnip row is the frequency below:
```

mpg	Freq.	Percent	Cum.
12	2	2.70	2.70
14	6	8.11	10.81
16	6	8.11	18.92
18	13	17.57	36.49
20	11	14.86	51.35
22	10	13.51	64.86
24	7	9.46	74.32
26	8	10.81	85.14
28	3	4.05	89.19
30	3	4.05	93.24
32	1	1.35	94.59
34	1	1.35	95.95
36	2	2.70	98.65
42	1	1.35	100.00
Total	74	100.00	

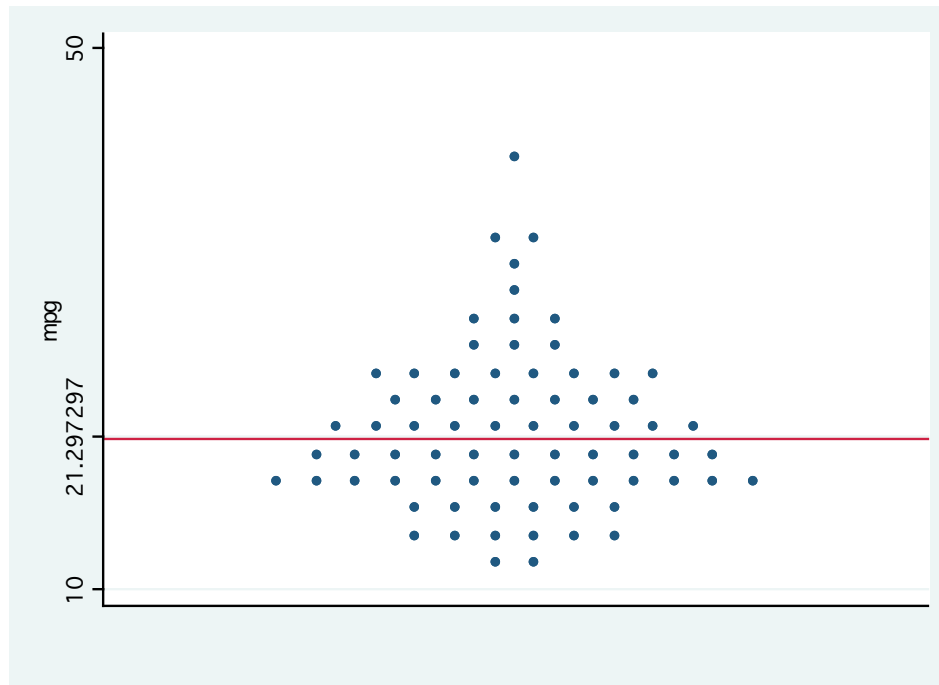
```
options: xlabel(-10 10) yline(21.2) ylabel( 21.2972972972973 10 50)
```

```
resolution: 2
```

```
truev: 1.00e+11
```

```
dflag = mean
```

```
yline: 21.2972972972973 (mean)
```



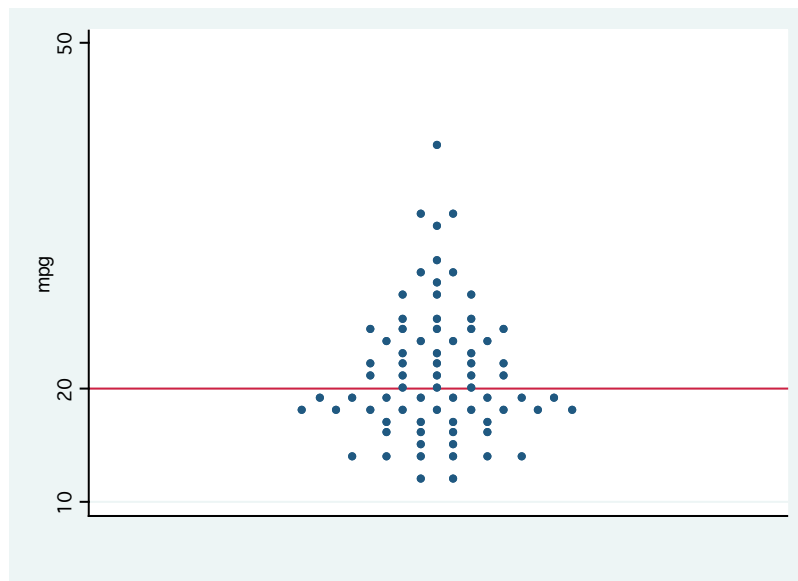
VIII. Draw yline at the median

```
turnip mpg , xlabel(-10 10) yline(median) ylabel(10 50) res(1)  
*same thing, but now draw a line at the median.
```

of observations per turnip row is the frequency below:

mpg	Freq.	Percent	Cum.
12	2	2.70	2.70
14	6	8.11	10.81
15	2	2.70	13.51
16	4	5.41	18.92
17	4	5.41	24.32
18	9	12.16	36.49
19	8	10.81	47.30
20	3	4.05	51.35
21	5	6.76	58.11
22	5	6.76	64.86
23	3	4.05	68.92
24	4	5.41	74.32
25	5	6.76	81.08
26	3	4.05	85.14
28	3	4.05	89.19
29	1	1.35	90.54
30	2	2.70	93.24
31	1	1.35	94.59
34	1	1.35	95.95
35	2	2.70	98.65
41	1	1.35	100.00
Total	74	100.00	

```
options: xlabel(-10 10) yline(20) ylabel( 20 10 50)  
resolution: 1  
truev: 1.00e+11  
dflag = median  
yline: 20 (median)
```



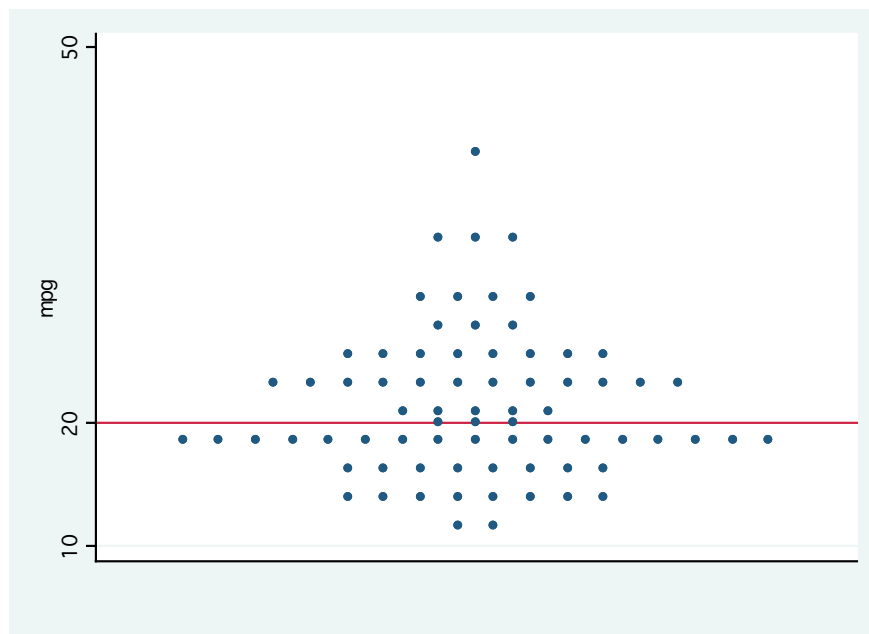
IX. Use the true value function (compare to graph VIII)

turnip mpg , xlabel(-10 10) yline(median) ylabel(10 50) t(20)
 *now repeat the graph. In this case, however, specify that only values truly = 20 get represented as a y=20 on the graph (ie, suspend any rounding at 20). Here, 3 observations are really at an mpg of 20. These had been rounded to 20.79 (see example VIII).

```
. turnip mpg , xlabel(-10 10) yline(median) ylabel(10 50) t(20)
(74 real changes made)
(40 real changes made)
# of observations per turnip row is the frequency below:
```

mpg	Freq.	Percent	Cum.
11.55	2	2.70	2.70
13.86	8	10.81	13.51
16.17	8	10.81	24.32
18.48	17	22.97	47.30
20	3	4.05	51.35
20.79	5	6.76	58.11
23.1	12	16.22	74.32
25.41	8	10.81	85.14
27.72	3	4.05	89.19
30.03	4	5.41	94.59
34.65	3	4.05	98.65
41.58	1	1.35	100.00
Total	74	100.00	

```
options: xlabel(-10 10) yline(20) ylabel( 20 10 50)
resolution: 2.31
truev: 20
dflag = median
yline: 20 (median)
```



X. Combine turnip graphs

*for details/options about cr graph see help @gr combine@

```
turnip mpg, ylabel(0(10)60)
graph save mpg.gph"
turnip turn, ylabel(0(10)60)
graph save turn.gph"
turnip trunk, ylabel(0(10)60)
graph save trunk.gph"
gr combine mpg.gph turn.gph trunk.gph, row(1)
```

