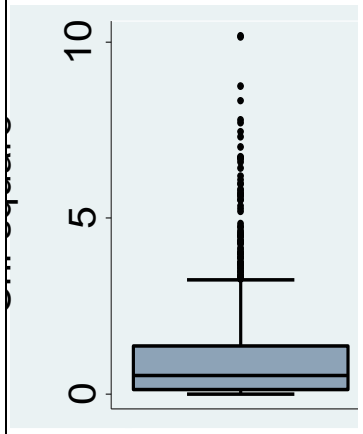
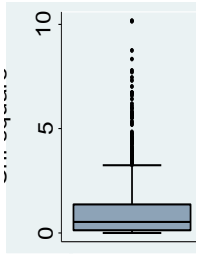
Biostat 202: Data visualization: Telling stories with data

Aaron Wolfe Scheffler,
Division of Biostatistics,
Department of Epidemiology and Biostatistics



Biostat 202



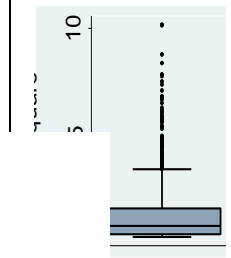
Outline for today

- Tables versus graphs

- Table principles
- Table practice

- Graph principles
- Graph types
- Graph practice

Study of heart disease patients



- Goal is to predict Major Adverse Cardiac Events (MACE) and mortality.
- Possible predictors: Left Ventricular Mass of the heart (LV mass) estimated during three heart phases and normalized by (divided by) body surface area (BSA) or weight. Also ejection fraction.

EUROPEAN SOCIETY OF CARDIOLOGY®

European Heart Journal – Cardiovascular Imaging (2017) **18**, 95–102
doi:10.1093/ehjci/jev357

Cardiac CT assessment of left ventricular mass in mid-diastasis and its prognostic value

Ran Klein[1]*, Emmanuelle S. Ametepe[2], Yeung Yam[3], Girish Dwivedi[3], and Benjamin J. Chow[3]

Table example – what is right/wrong?

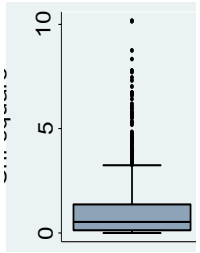


Table 3 ROC analysis results for predictors of MACE and all-cause mortality

Parameter	Optimal threshold			ROC AUC		
	M	F	M + F	M	F	M + F
$LVMI_{BSA}$ (g/m^2)						
MD	79.2[a]	63.8[a]	81.7	0.572	0.744	0.654
ED	N.A.	72.4	83.3	0.561	0.672	0.626
ES	N.A.	72.3	79.4	0.544	0.732	0.631
$LVMI_{Weight}$ (g/kg)						
MD	2.13	1.72	1.97	0.611	0.754	0.685
ED	2.09	1.75	1.98	0.616	0.675	0.660
ES	N.A.	1.76	1.96	0.583	0.742	0.662
LVEF%	62.7	67.4	61.5	0.654	0.670	0.664

No significant differences in AUC were detected between phases ($P > 0.05$ using a Bonferroni adjustment for multiple comparisons.
ROC AUC, receiver-operator characteristics–area under curve; N.A., not available due to not well-defined optimal threshold; LV, left ventricle.
[a]Thresholds from Mao *et al.* (2013).

European Heart Journal - Cardiovascular Imaging (2017) 18, 95-102
doi:10.1093/ehjci/jev357

Table example – focus on right side

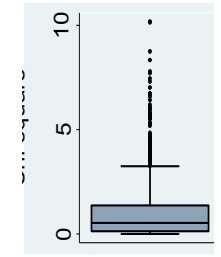
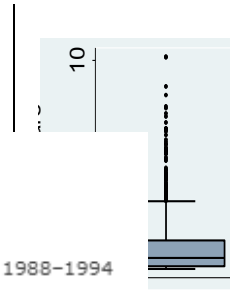


Table 3 ROC analysis results for predictors of MACE and all-cause mortality

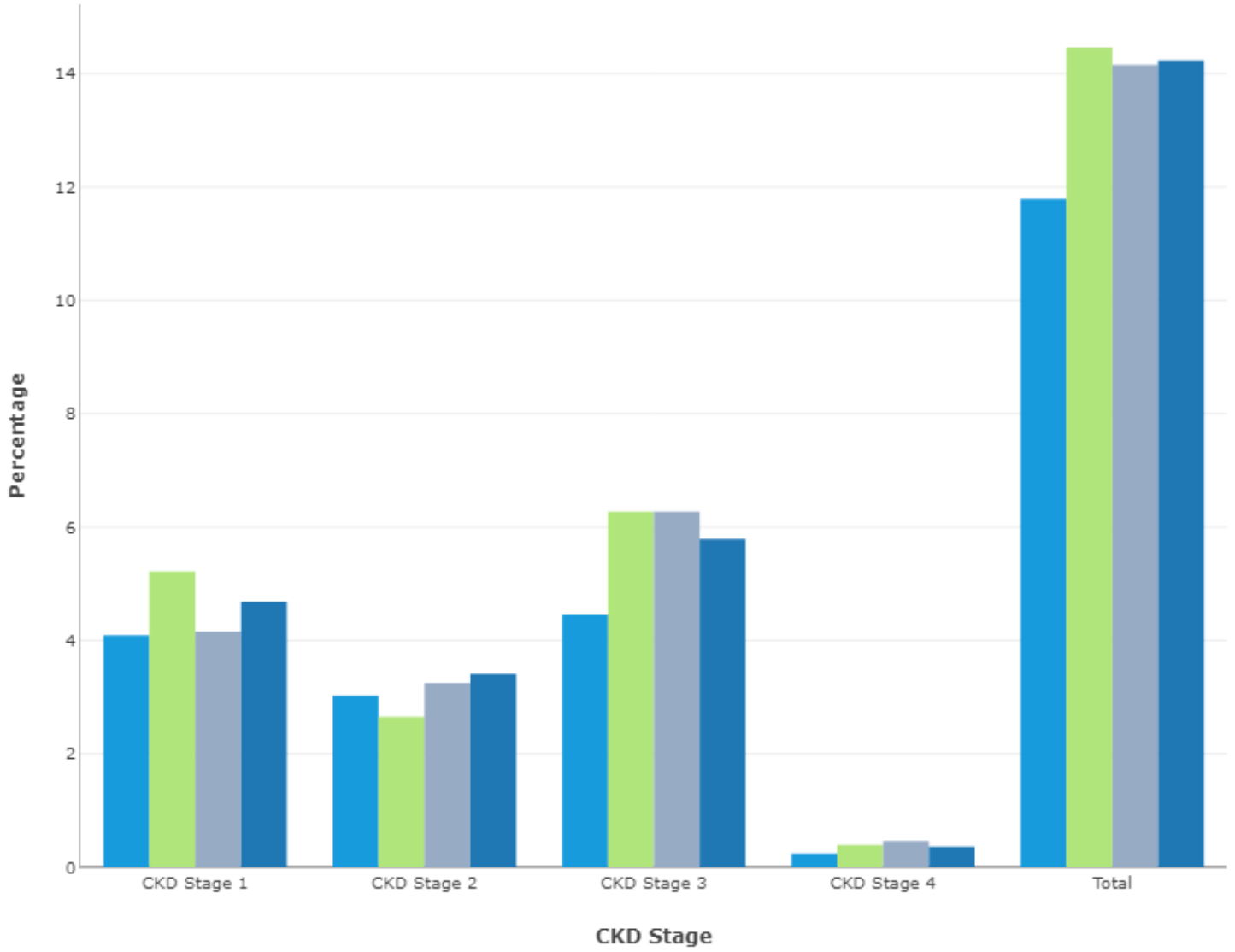
Parameter	ROC AUC		
	M	F	M + F
$LVMI_{BSA}$ (g/m^2)			
MD	0.572	0.744	0.654
ED	0.561	0.672	0.626
ES	0.544	0.732	0.631
$LVMI_{Weight}$ (g/kg)			
MD	0.611	0.754	0.685
ED	0.616	0.675	0.660
ES	0.583	0.742	0.662
LVEF%	0.654	0.670	0.664

No significant differences in AUC were detected between phases ($P > 0.05$ using a Bonferroni adjustment for multiple comparisons.
ROC AUC, receiver-operator characteristics–area under curve; N.A., not available due to not well-defined optimal threshold; LV, left ventricle.
[a]Thresholds from Mao *et al.* (2013).

Graph example – what is right/wrong?

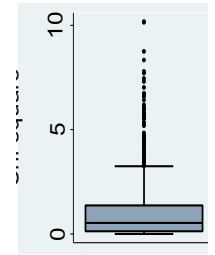


Prevalence of CKD by CKD Stage and Year, 1988-1994 to 2015-2016
National Health and Nutrition Examination Survey



Centers for Disease Control and Prevention. Chronic Kidney Disease Surveillance System—United States. website. https://nccd.cdc.gov/ckd

Same data in a table – what is right/wrong? Better/worse?

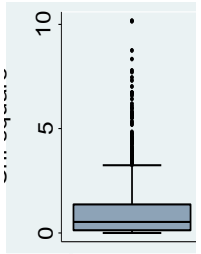


Prevalence of CKD by CKD Stage and Year, 1988-1994 to 2015-2016[a,b]
National Health and Nutrition Examination Survey

Year	CKD Stage 1	CKD Stage 2	CKD Stage 3	CKD Stage 4	Total
1988–1994	4.1% (CI 3.6-4.6)	3.0% (CI 2.6-3.4)	4.5% (CI 3.9-5.0)	.2% (CI .2-.3)	11.8% (CI 11.0-12.6)
1999–2006	5.2% (CI 4.7-5.8)	2.7% (CI 2.3-3.0)	6.3% (CI 5.7-6.7)	.4% (CI .3-.5)	14.5% (CI 13.7-15.3)
2007–2014	4.2% (CI 3.8-4.5)	3.3% (CI 2.9-3.6)	6.3% (CI 5.8-6.7)	.5% (CI .4-.6)	14.2% (CI 13.5-14.8)
2015–2016	4.7% (CI 3.9-5.5)	3.4% (CI 2.3-4.6)	5.8% (CI 4.6-7.0)	.4% (CI .3-.5)	14.2% (CI 12.5-16.0)

Notes:
95% confidence intervals, when available, are shown in parentheses.

Outline for today

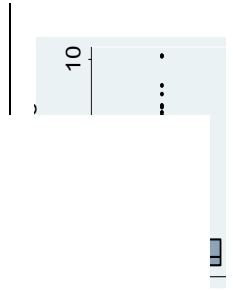


- Tables versus graphs

- Table principles
- Table practice

- Graph principles
- Graph types
- Graph practice

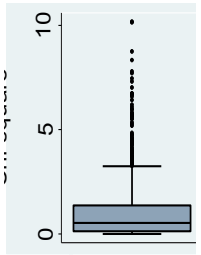
Tables vs graphs



Use **TABLES** when

- The user needs to look up specific values
- You want to enable individual quantitative comparisons
- You need to convey results to high precision
- You want to convey both individual and summary information
- You want to display items that are measured on very different scales. Workarounds: left vs right scales; log scales; multiple panels.

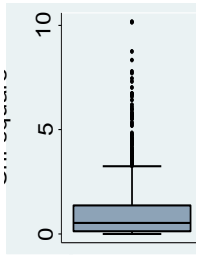
Tables vs graphs

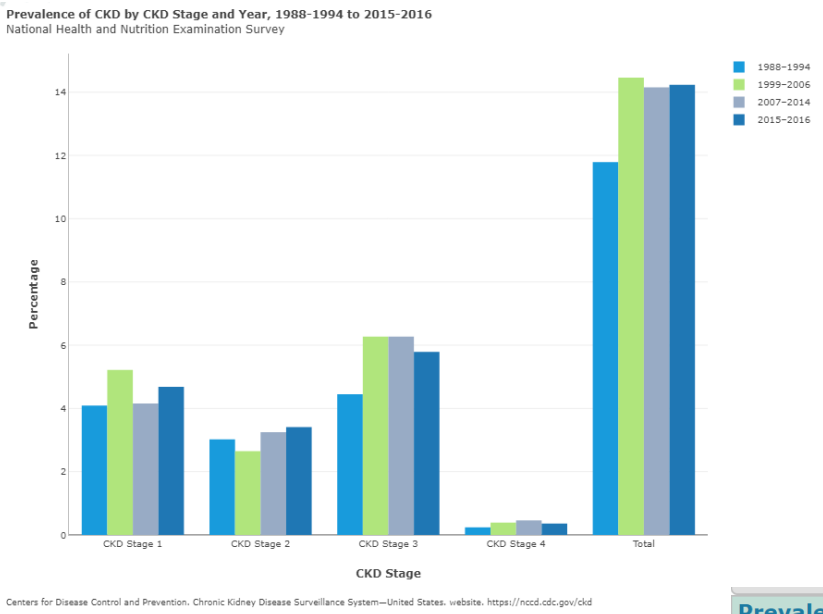


Use **GRAPHS** when

- You want to convey shapes
- You want to convey patterns
- You want to convey trends
- You want to convey quantitative ideas quickly
- *Corollary: You should minimize tables in scientific presentations in favor of graphs!*

Chronic Kidney Disease





Centers for Disease Control and Prevention. Chronic Kidney Disease Surveillance System—United States. website. https://nccd.cdc.gov/ckd

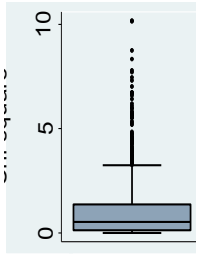
Which when?

Prevalence of CKD by CKD Stage and Year, 1988-1994 to 2015-2016[a,b]
National Health and Nutrition Examination Survey

Year	CKD Stage 1	CKD Stage 2	CKD Stage 3	CKD Stage 4	Total
1988–1994	4.1% (CI 3.6-4.6)	3.0% (CI 2.6-3.4)	4.5% (CI 3.9-5.0)	.2% (CI .2-.3)	11.8% (CI 11.0-12.6)
1999–2006	5.2% (CI 4.7-5.8)	2.7% (CI 2.3-3.0)	6.3% (CI 5.7-6.7)	.4% (CI .3-.5)	14.5% (CI 13.7-15.3)
2007–2014	4.2% (CI 3.8-4.5)	3.3% (CI 2.9-3.6)	6.3% (CI 5.8-6.7)	.5% (CI .4-.6)	14.2% (CI 13.5-14.8)
2015–2016	4.7% (CI 3.9-5.5)	3.4% (CI 2.3-4.6)	5.8% (CI 4.6-7.0)	.4% (CI .3-.5)	14.2% (CI 12.5-16.0)

Notes:
95% confidence intervals, when available, are shown in parentheses.

Outline for today

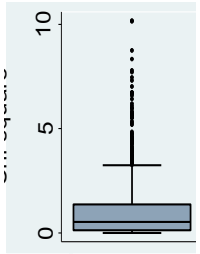


- Tables versus graphs

- Table principles
- Table practice

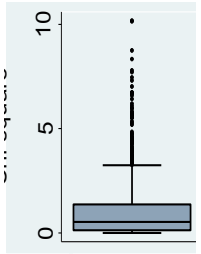
- Graph principles
- Graph types
- Graph practice

Table principles



- Should be <u>self-explanatory!!!</u>
- It is easier to compare numbers in columns.
- Round aggressively, preferably to two effective digits (ie where the numbers differ). *Exception: you need to convey precise results.*
- Order rows and/or columns purposefully to convey information. Compare across columns
- Lightly use visual elements to ease comparisons.

Outline for today



- Tables versus graphs

- Table principles
- Table practice

- Graph principles
- Graph types
- Graph practice

Table example – focus on right side

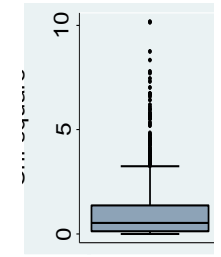
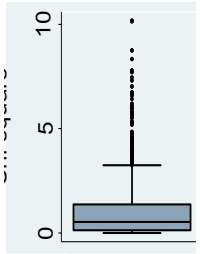


Table 3 ROC analysis results for predictors of MACE and all-cause mortality

Parameter	ROC AUC		
	M	F	M + F
$LVMI_{BSA}$ (g/m^2)			
MD	0.572	0.744	0.654
ED	0.561	0.672	0.626
ES	0.544	0.732	0.631
$LVMI_{Weight}$ (g/kg)			
MD	0.611	0.754	0.685
ED	0.616	0.675	0.660
ES	0.583	0.742	0.662
LVEF%	0.654	0.670	0.664

No significant differences in AUC were detected between phases ($P > 0.05$ using a Bonferroni adjustment for multiple comparisons.
ROC AUC, receiver-operator characteristics–area under curve; N.A., not available due to not well-defined optimal threshold; LV, left ventricle.
[a]Thresholds from Mao *et al.* (2013).

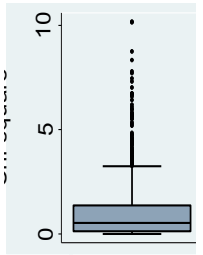
Improving the cardiac prediction table



- Improve table with
  - Explanation
  - Rounded to 2 digits
  - Light visuals
  - Order columns to put ‘M+F’ on the left
- What’s good? has primary comparisons in columns, lightly uses visual elements, and is ordered purposefully.

Table 3: Area under the ROC curve (AUROC) analysis for different predictors of MACE and all-cause mortality by phase of the cardiac cycle for all participants, males, and females. Prediction was better for females than for males and differed minimally by phase. LVMI normalized by weight performed better than normalization by body surface area and was slightly better than LVEF.

	AUROC		
Predictor	All	Males	Females
$LVMI_{BSA}$ (g/m^2)			
MD	.65	.57	.74
ED	.63	.56	.67
ES	.63	.54	.73
$LVMI_{Weight}$ (g/kg)			
MD	.69	.61	.75
ED	.66	.62	.68
ES	.66	.58	.74
LVEF%			
	.66	.65	.67



From: **Trends in Cardiometabolic Mortality in the United States, 1999-2017**

JAMA. 2019;322(8):780-782. doi:10.1001/jama.2019.9161

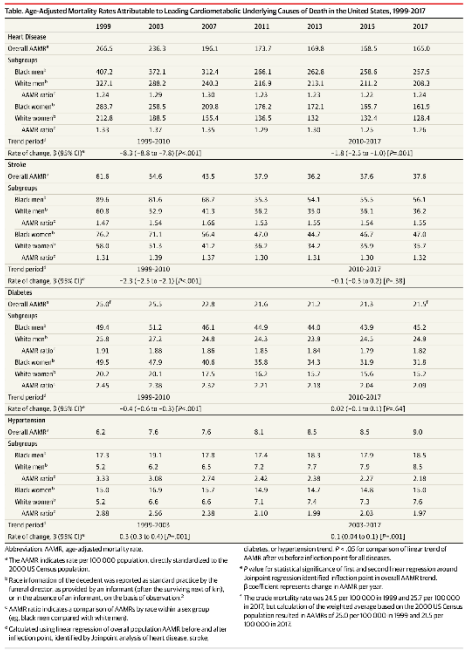


Table. Age-Adjusted Mortality Rates Attributable to Leading Cardiometabolic Underlying Causes of Death in the United States, 1999-2017

	1999	2003	2007	2011	2013	2015	2017
Heart Disease							
Overall AAMR[a]	265.5	236.3	196.1	173.7	169.8	168.5	165.0
Subgroups							
Black men[b]	407.2	372.1	312.4	286.1	262.8	258.6	257.5
White men[b]	327.1	288.2	240.3	216.9	213.1	211.2	208.3
AAMR ratio[c]	1.24	1.29	1.30	1.32	1.23	1.22	1.24
Black women[b]	283.7	258.5	209.8	178.2	172.1	165.7	161.5
White women[b]	212.8	188.5	155.4	136.5	132	132.4	128.4
AAMR ratio[c]	1.33	1.37	1.35	1.29	1.30	1.25	1.26
Trend period[d]		1999-2010				2010-2017	
Rate of change, % (95% CI)[e]		−6.3 (−6.8 to −7.9) [P<.001]				−1.8 (−2.5 to −1.0) [P=.001]	
Stroke							
Overall AAMR[a]	61.8	54.6	43.5	37.9	36.2	37.6	37.8
Subgroups							
Black men[b]	89.6	83.6	66.7	55.5	54.1	55.5	56.1
White men[b]	60.8	52.9	41.3	36.2	35.0	36.1	36.2
AAMR ratio[c]	1.47	1.54	1.66	1.53	1.55	1.54	1.55
Black women[b]	76.2	71.1	56.4	47.0	44.7	46.7	47.0
White women[b]	58.0	51.5	41.2	36.2	34.2	35.9	35.7
AAMR ratio[c]	1.31	1.39	1.37	1.30	1.31	1.30	1.32
Trend period[d]		1999-2010				2010-2017	
Rate of change, % (95% CI)[e]		−2.3 (−2.5 to −2.1) [P<.001]				−0.1 (−0.5 to 0.3) [P=.58]	
Diabetes							
Overall AAMR[a]	26.8[f]	25.5	22.8	21.6	21.2	21.6	21.9[f]
Subgroups							
Black men[b]	49.4	51.2	46.1	44.9	44.0	43.9	45.2
White men[b]	24.8	23.2	24.8	24.3	23.9	24.3	24.9
AAMR ratio	1.91	1.88	1.86	1.85	1.84	1.79	1.82
Black women[b]	49.5	47.9	40.6	35.8	34.3	31.9	31.8
White women[b]	20.2	20.1	17.5	16.2	15.7	15.6	15.2
AAMR ratio	2.45	2.38	2.32	2.21	2.18	2.04	2.09
Trend period[d]		1999-2010				2010-2017	
Rate of change, % (95% CI)[e]		−0.4 (−0.6 to −0.3) [P<.001]				0.02 (−0.1 to 0.1) [P=.64]	
Hypertension							
Overall AAMR[a]	6.2	7.6	7.6	8.1	8.5	8.5	9.0
Subgroups							
Black men[b]	17.3	19.1	17.8	17.4	18.3	17.9	18.5
White men[b]	5.2	6.2	6.5	7.2	7.7	7.9	8.5
AAMR ratio[c]	3.33	3.08	2.74	2.42	2.38	2.27	2.18
Black women[b]	13.8	16.9	13.7	14.9	14.7	14.8	15.8
White women[b]	5.2	6.6	6.6	7.1	7.4	7.3	7.6
AAMR ratio[c]	2.88	2.56	2.38	2.10	1.99	2.03	1.97
Trend period[d]		1999-2003				2003-2017	
Rate of change, % (95% CI)[e]		0.3 (0.3 to 0.4) [P<.001]				0.1 (0.04 to 0.1) [P=.001]	

Abbreviation: AAMR, age-adjusted mortality rate.

[a] The AAMR indicates rate per 100 000 population, directly standardized to the 2000 US Census population.

[b] Race information of the decedent was reported as ascertained per decedent's funeral director, as provided by an informant (often the surviving next of kin), or in the absence of an informant, on the basis of observation.[3]

[c] AAMR ratio indicates a comparison of AAMRs by race within a sex group (eg, Black men compared with White men).

[d] Calculated using linear regression of overall population AAMR before and after inflection point, identified by Joinpoint analysis of heart disease, stroke, diabetes, or hypertension trend. P < .05 for comparison of linear trend of AAMR after vs before inflection point for all diseases.

[e] P value for statistical significance of first and second linear regression around Joinpoint regression identified inflection point in overall AAMR trend. (Coefficient represents change in AAMR per year.

[f] The crude mortality rate was 24.5 per 100 000 in 1999 and 25.7 per 100 000 in 2017, but calculation of the weighted average based on the 2000 US Census population resulted in AAMRs of 25.9 per 100 000 in 1999 and 26.5 per 100 000 in 2017.

Table Title:

Age-Adjusted Mortality Rates Attributable to Leading Cardiometabolic Underlying Causes of Death in the United States, 1999-2017

Improving the mortality trend table

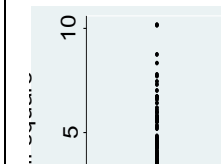


Table. Age-Adjusted Mortality Rates Attributable to Leading Cardiometabolic Underlying Causes of Death in the United States, 1999-2017

	1999	2003	2007	2011	2013	2015	2017
Heart Disease							
Overall AAMR[a]	266.5	236.3	196.1	173.7	169.8	168.5	165.0
Subgroups							
Black men[b]	407.2	372.1	312.4	266.1	262.8	258.6	257.5
White men[b]	327.1	288.2	240.3	216.9	213.1	211.2	208.3
AAMR ratio[c]	1.24	1.29	1.30	1.23	1.23	1.22	1.24
Black women[b]	283.7	258.5	209.8	176.2	172.1	165.7	161.9
White women[b]	212.8	188.5	155.4	136.5	132	132.4	128.4
AAMR ratio[c]	1.33	1.37	1.35	1.29	1.30	1.25	1.26
Trend period[d]	1999-2010				2010-2017		
Rate of change, β (95% CI)[e]	−8.3 (−8.8 to −7.8) [*P*<.001]				−1.8 (−2.5 to −1.0) [*P*=.001]		

Abbreviation: AAMR, age-adjusted mortality rate.

[a] The AAMR indicates rate per 100 000 population, directly standardized to the 2000 US Census population.

[b] Race information of the decedent was reported as standard practice by the funeral director, as provided by an informant (often the surviving next of kin), or in the absence of an informant, on the basis of observation.[2]

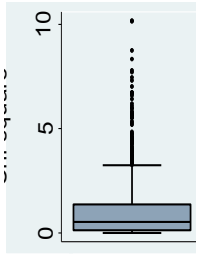
[c] AAMR ratio indicates a comparison of AAMRs by race within a sex group (eg, black men compared with white men).

[d] Calculated using linear regression of overall population AAMR before and after inflection point, identified by Joinpoint analysis of heart disease, stroke, diabetes, or hypertension trend. $P < .05$ for comparison of linear trend of AAMR after vs before inflection point for all diseases.

[e] *P* value for statistical significance of first and second linear regression around Joinpoint regression identified inflection point in overall AAMR trend. β coefficient represents change in AAMR per year.

[f] The crude mortality rate was 24.5 per 100 000 in 1999 and 25.7 per 100 000 in 2017, but calculation of the weighted average based on the 2000 US Census population resulted in AAMRs of 25.0 per 100 000 in 1999 and 21.5 per 100 000 in 2017.

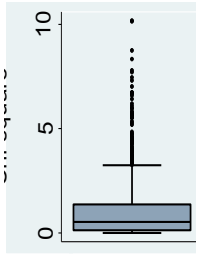
Improving the mortality trend table



- Improve table with
  - Rounded to 2 effective digits – how?
  - Light visuals instead of overall shading
  - Main comparison in columns?

- Perhaps better as a graph. Why a table?
  - Trying to get a lot of information in here

Outline for today

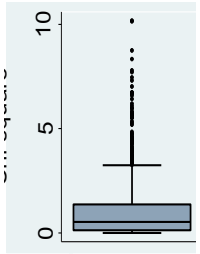


- Tables versus graphs

- Table principles
- Table practice

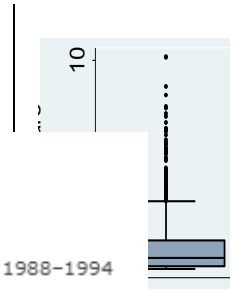
- Graph principles
- Graph types
- Graph practice

Graph principles

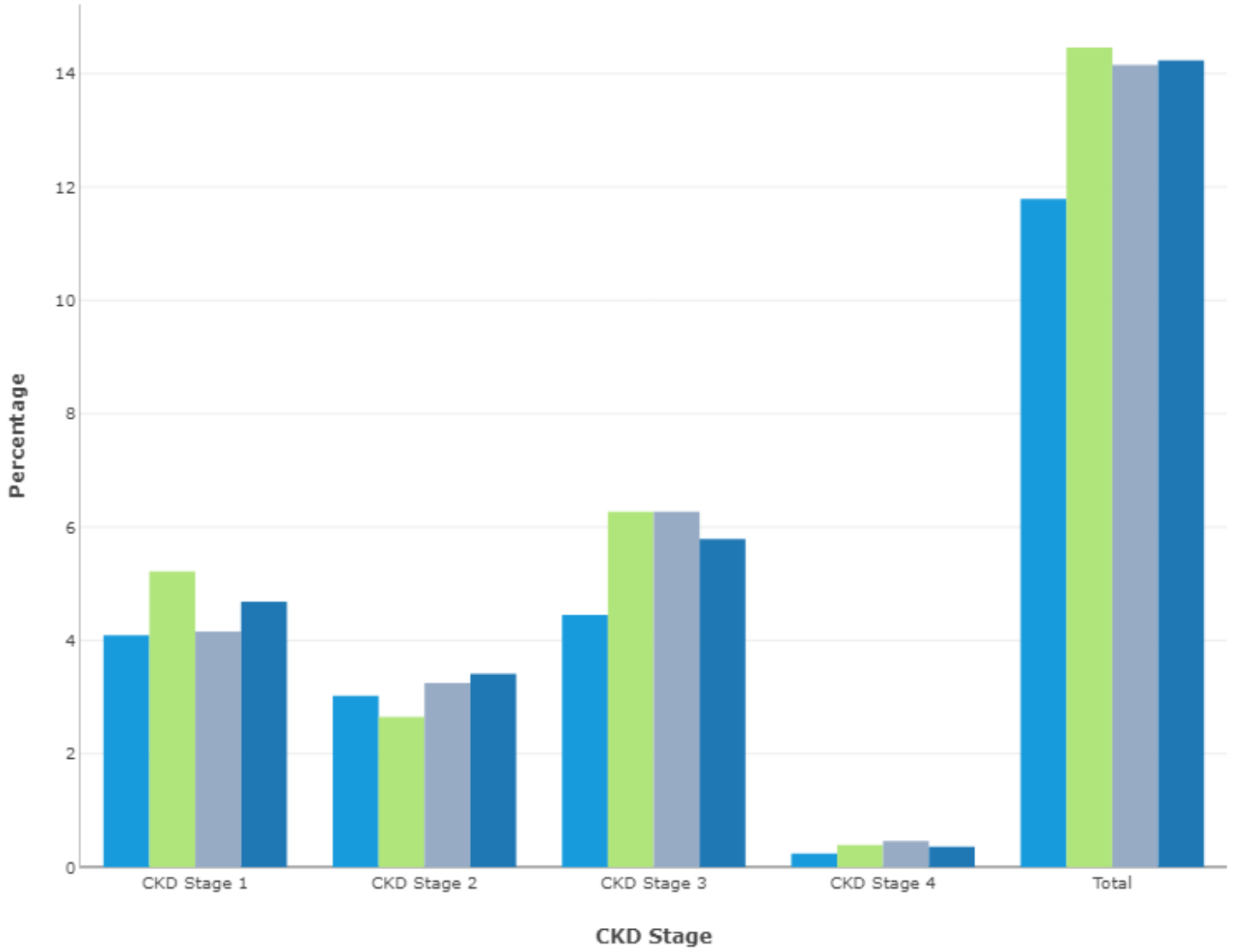


- Should be self-explanatory.
- Aim for interocular traumatic impact. Use the ideas below to convey your message clearly.
- Choose a graph type to emphasize your desired message.
- **Maximize the data to ink ratio**.
- Minimize the impact of non-data ink.
- Use colors, marker symbols and line types purposefully (Excel is ☹).

CKD graph: how to improve?

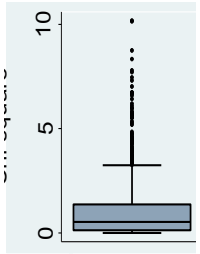


Prevalence of CKD by CKD Stage and Year, 1988-1994 to 2015-2016
National Health and Nutrition Examination Survey



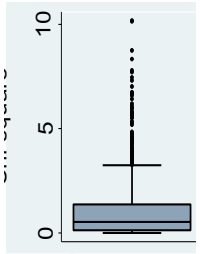
Centers for Disease Control and Prevention. Chronic Kidney Disease Surveillance System—United States. website. https://nccd.cdc.gov/ckd

How could the CKD graph be improved?



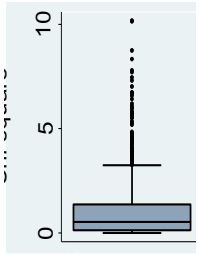
- Explanation
- Drop total columns
- Change scale
- Change color scheme
- *Try your hand at it in the homework!*

Graphs can convey data by



- Points – location in space
- Lines – trends (also points+lines)
- Bars – by height (also points+bars)
- Colors
- Point types
- Point size
- Line types
- Bar or symbol fill (color or shading)

Outline for today

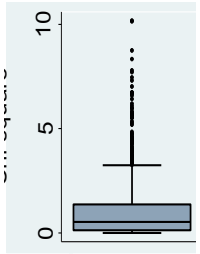


- Tables versus graphs

- Table principles
- Table practice

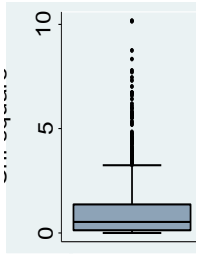
- Graph principles
- Graph types
- Graph practice

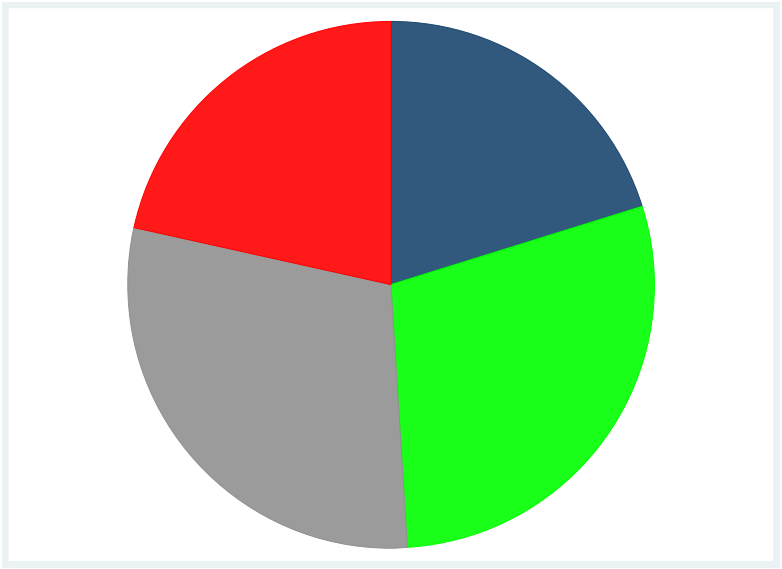
Graph types

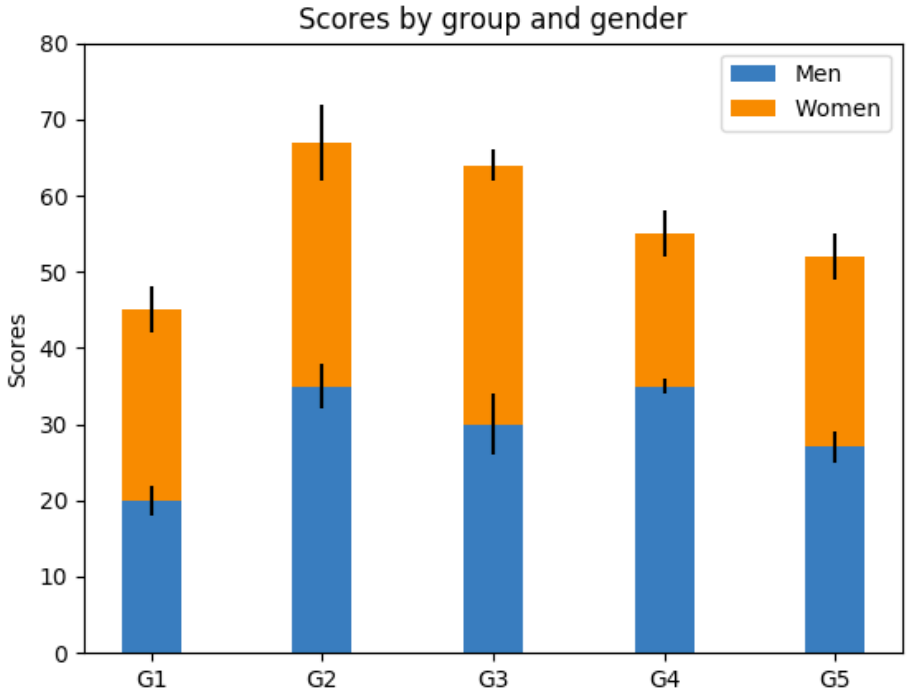


- Points – scatter plot for two numeric variables
- Lines – trend for numeric over time
- Points and lines – connected plot, smoothed plot, estimate and error bars
- Bars – bar chart for numeric by categorical; histogram for frequency versus numerical value to display shape.
- Bars and lines – bar chart with error bars; boxplot for distributional shape.

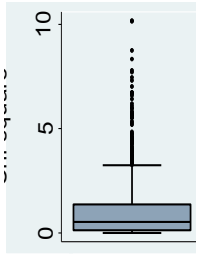
Avoid:
pie charts and stacked bar charts



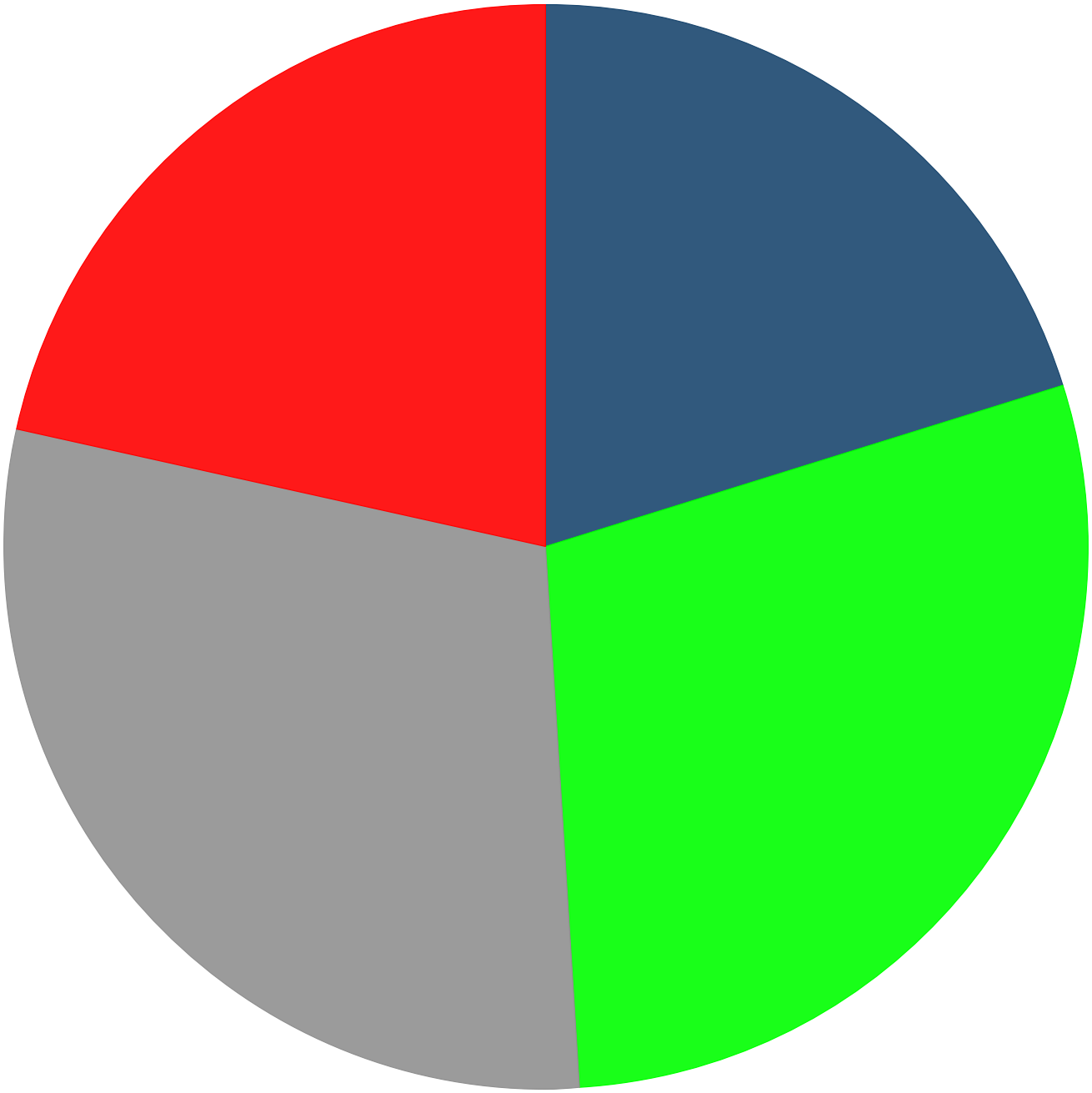


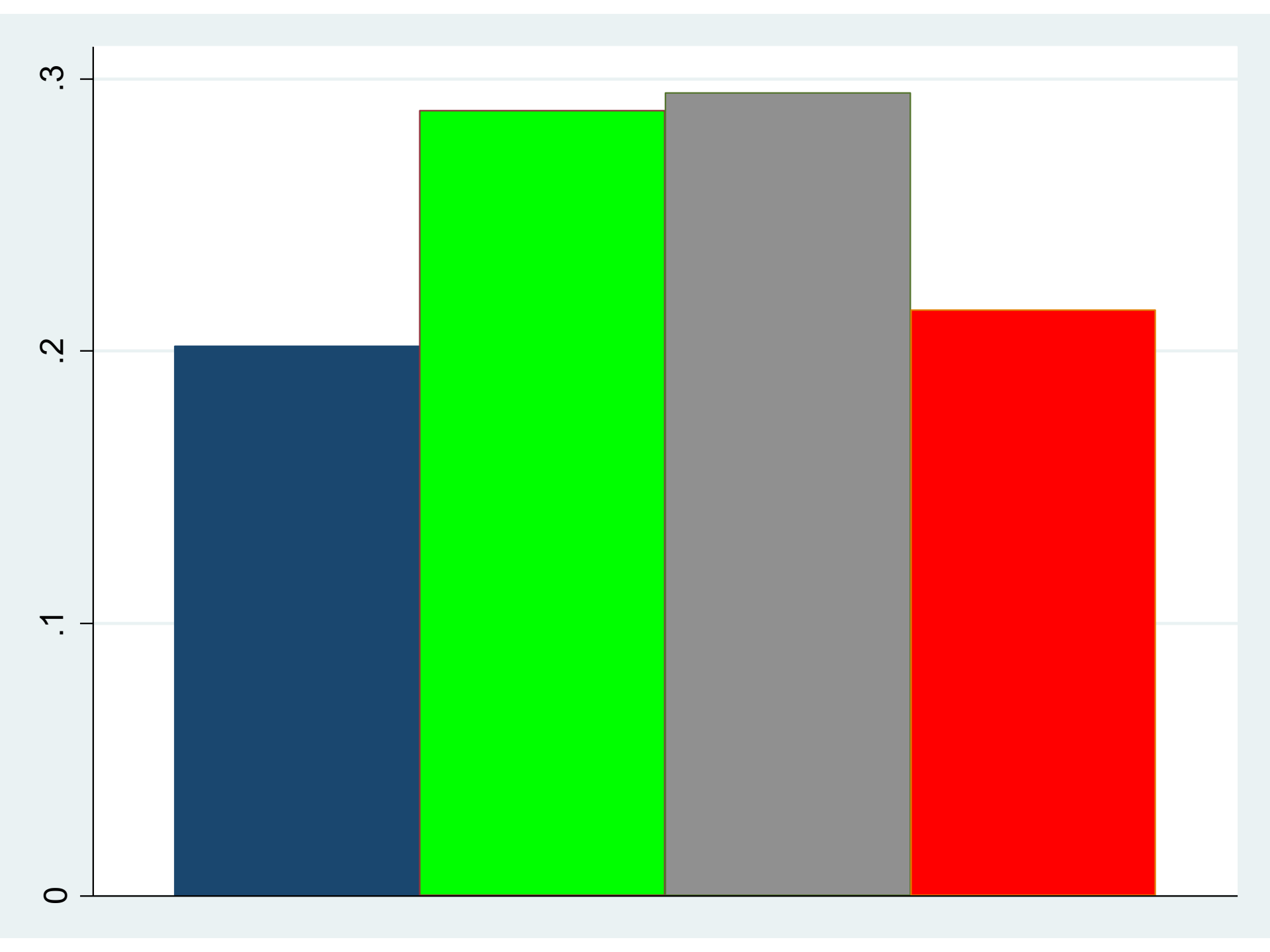


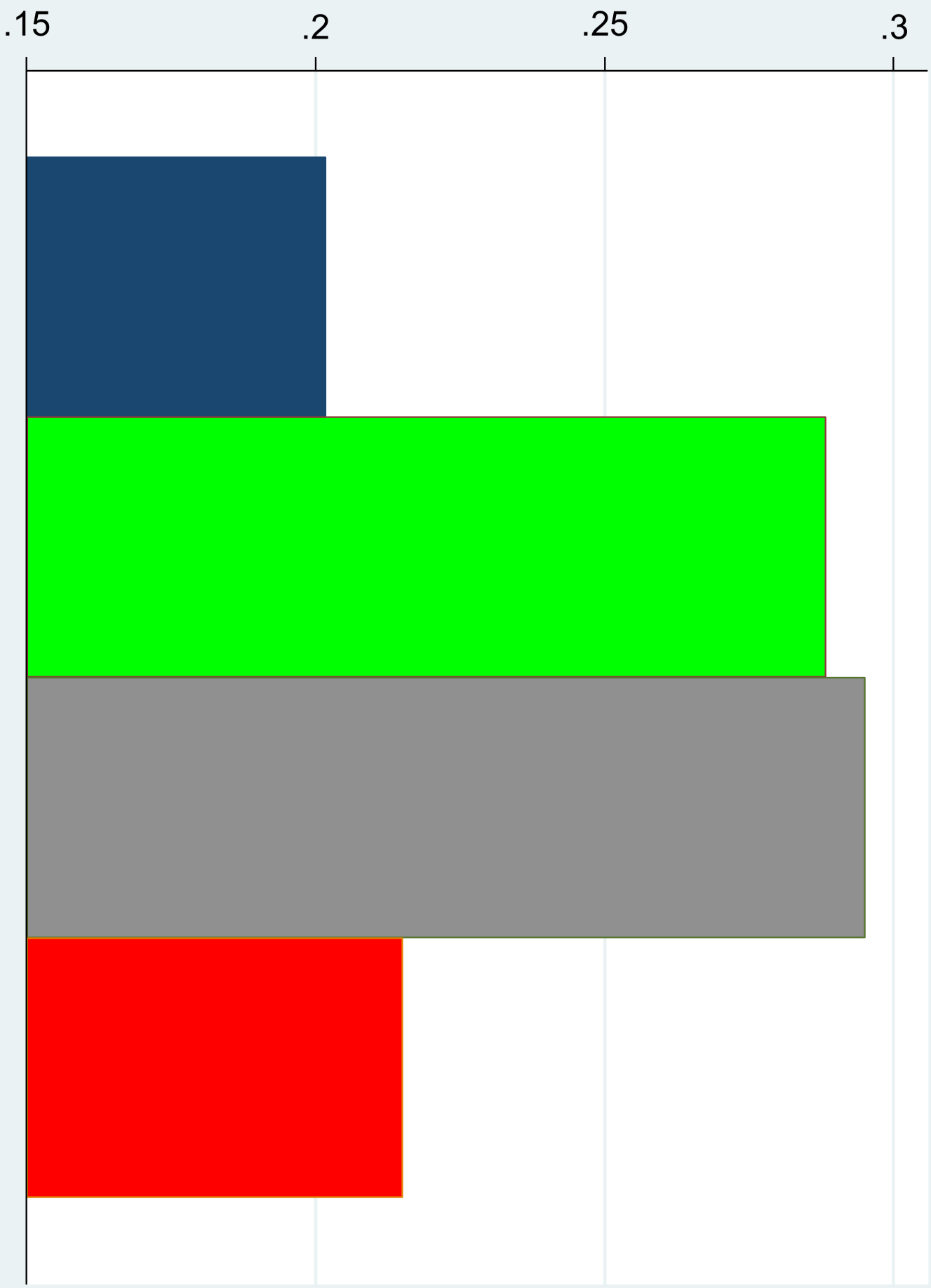
Avoid:
pie charts and stacked bar charts



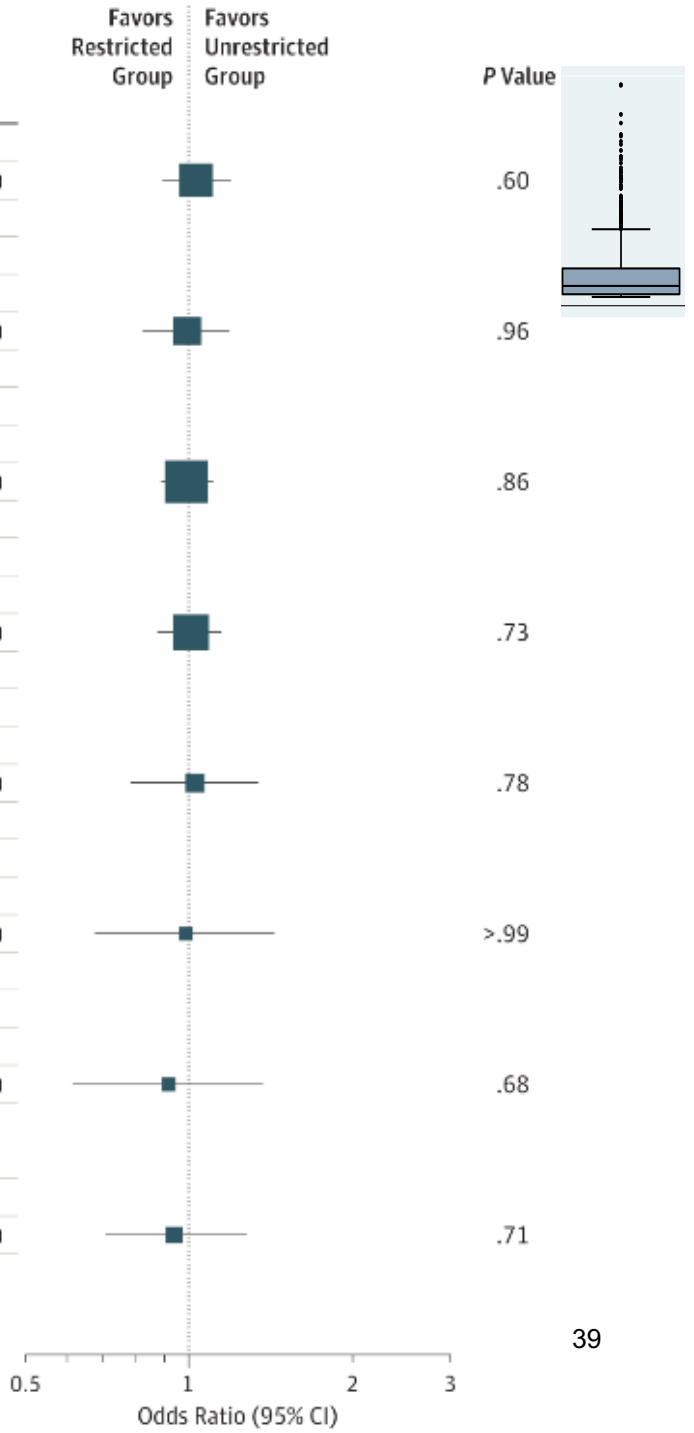
- Comparing the areas is a difficult visual perception task.
- We are really bad at comparing relative areas that are represented in these pie charts and stacked bar chart.
- Wanna bet? Let's see.



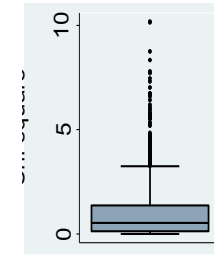

0
.1
.2
.3


.15
.2
.25
.3

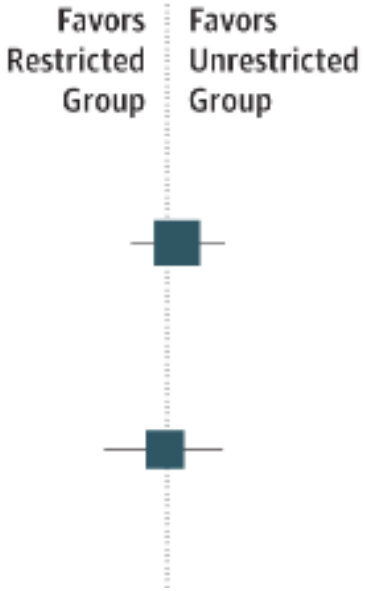
Practice Setting	No. of Order Sessions: Restricted Group[a]	No. of Order Sessions: Unrestricted Group[b]	Odds Ratio[c] (95% CI)	*P* Value
Overall				
Wrong-patient order sessions per 100 000	90.7	88.0	1.03 (0.90-1.20)	.60
Wrong-patient order sessions	1980	2026		
Total order sessions	2 183 365	2 303 266		
Emergency department				
Wrong-patient order sessions per 100 000	157.8	161.3	1.00 (0.83-1.20)	.96
Wrong-patient order sessions	560	576		
Total order sessions	354 882	357 047		
Inpatient				
Wrong-patient order sessions per 100 000	185.6	185.1	0.99 (0.89-1.11)	.86
Wrong-patient order sessions	1324	1340		
Total order sessions	713 417	723 746		
Medical/surgical				
Wrong-patient order sessions per 100 000	187.9	187.8	0.98 (0.85-1.12)	.73
Wrong-patient order sessions	940	879		
Total order sessions	500 338	467 941		
Critical care				
Wrong-patient order sessions per 100 000	247.8	258.2	1.04 (0.79-1.36)	.78
Wrong-patient order sessions	161	216		
Total order sessions	64 979	83 662		
Pediatrics				
Wrong-patient order sessions per 100 000	122.5	139.2	1.00 (0.68-1.46)	>.99
Wrong-patient order sessions	65	111		
Total order sessions	53 074	79 726		
Obstetrics				
Wrong-patient order sessions per 100 000	201.1	197.1	0.92 (0.61-1.38)	.68
Wrong-patient order sessions	85	78		
Total order sessions	42 272	39 575		
Outpatient				
Wrong-patient order sessions per 100 000	7.9	8.2	0.94 (0.70-1.28)	.71
Wrong-patient order sessions	86	97		
Total order sessions	1 082 855	1 176 344		



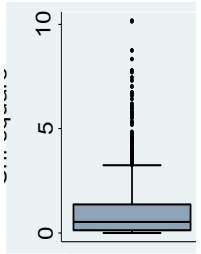
Having your cake and eating it too



Practice Setting	No. of Order Sessions		Odds Ratio[c] (95% CI)
	Restricted Group[a]	Unrestricted Group[b]	
Overall			
Wrong-patient order sessions per 100 000	90.7	88.0	1.03 (0.90-1.20)
Wrong-patient order sessions	1980	2026	
Total order sessions	2 183 365	2 303 266	
Emergency department			
Wrong-patient order sessions per 100 000	157.8	161.3	1.00 (0.83-1.20)
Wrong-patient order sessions	560	576	
Total order sessions	354 882	357 047	

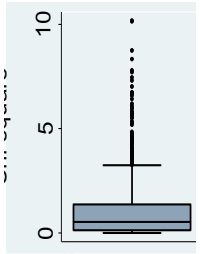


Graph types - Two at a time



- Two numerical variables
  - Scatterplot
- Numerical and Categorical
  - Bar chart
  - Box plot
- Two categorical
  - Table

Outline for today

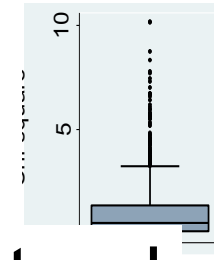


- Tables versus graphs

- Table principles
- Table practice

- Graph principles
- Graph types
- Graph practice

Ex: HPV vaccine uptake in Netherlands

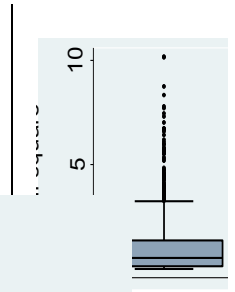


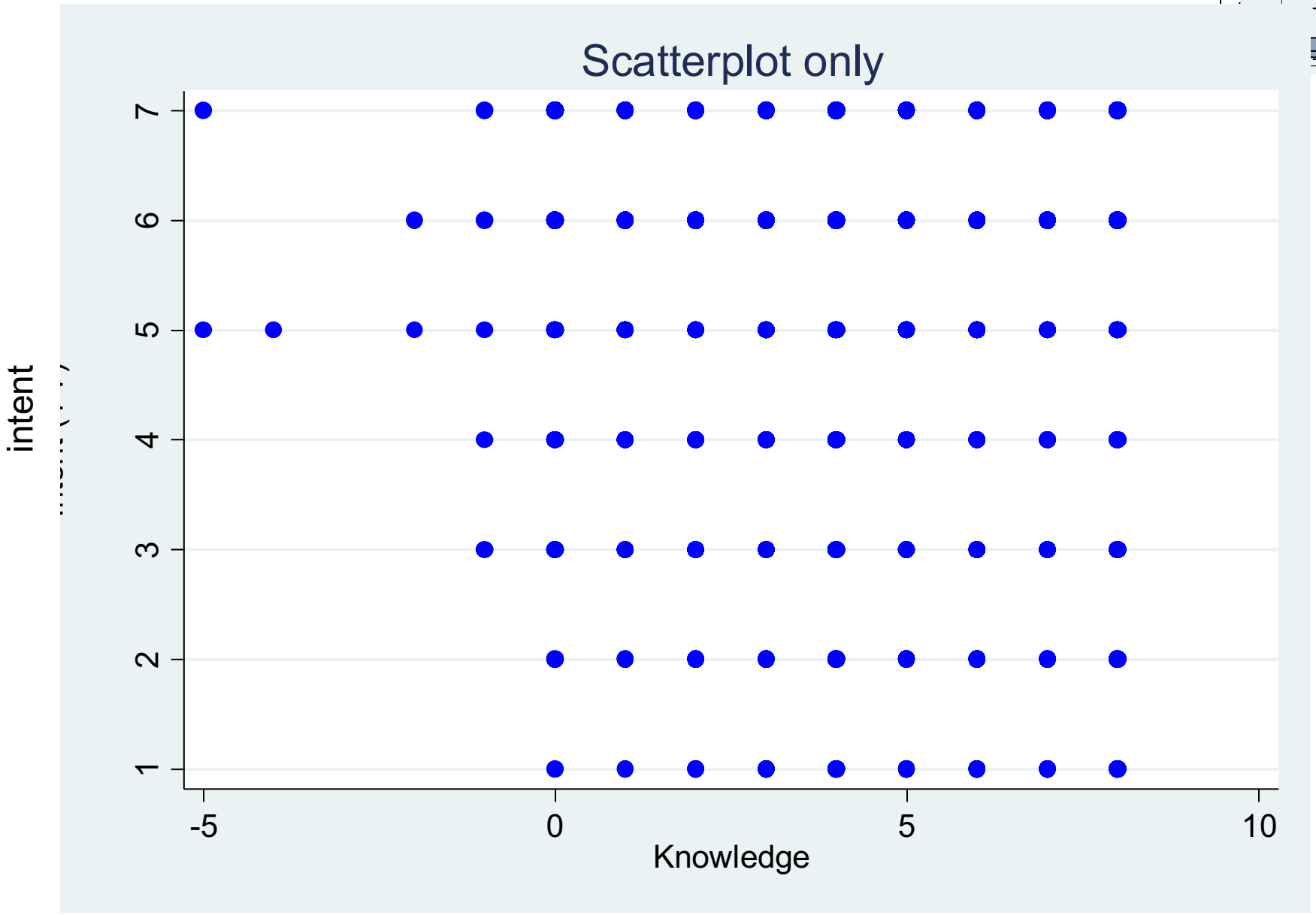
Pot *et al* (*Prev Med*, 2017) wanted to understand which variables were associated with intent for mothers to get their 12 year-old daughters vaccinated. Intention is measured on a scale from 1 (definitely not) to 7 (definitely yes). 8,062 responses.

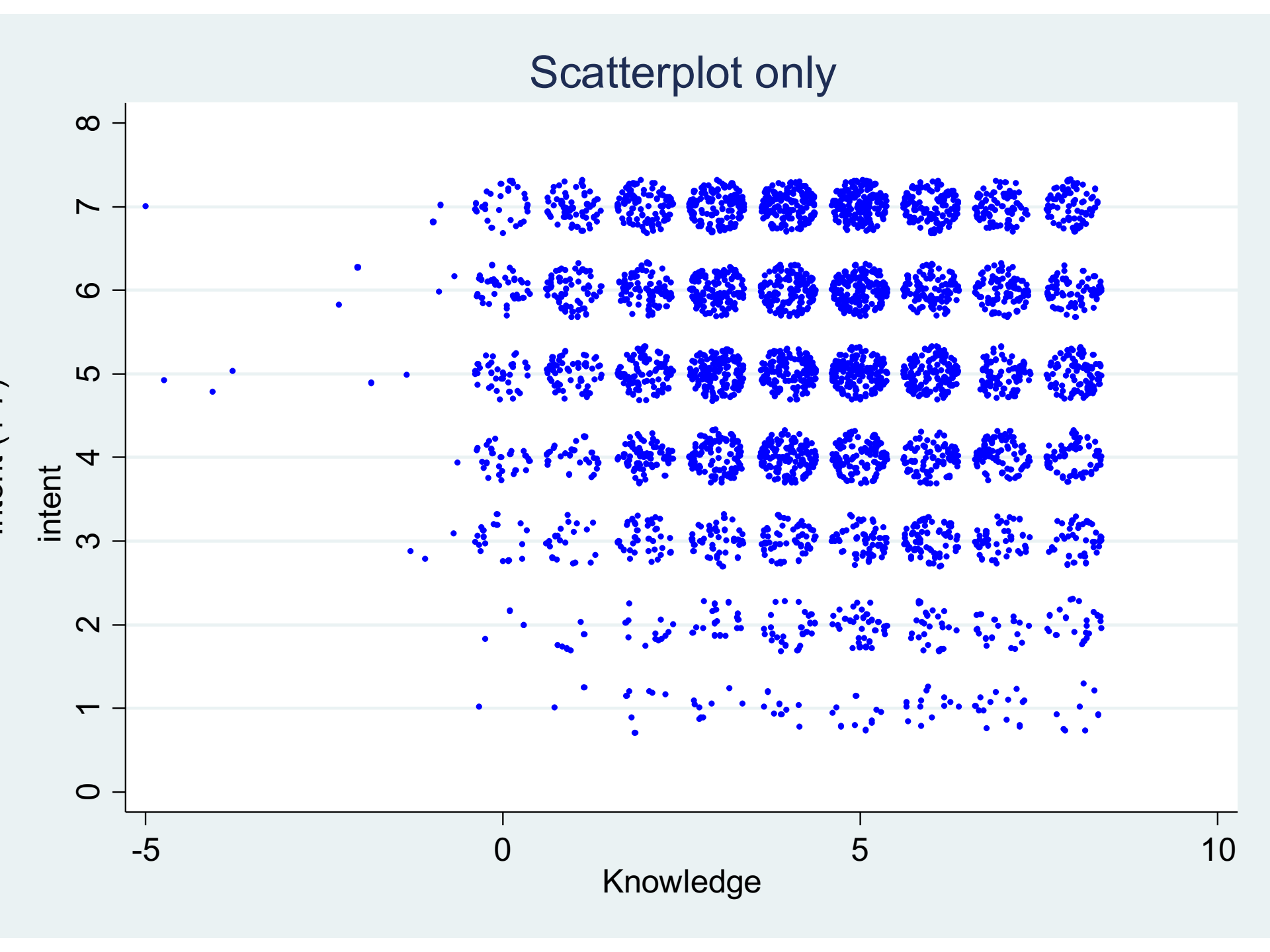
One of those predictors was knowledge about the HPV vaccine. Knowledge was assessed as the sum across 8 questions with -1 for wrong, +1 for correct (so a range of -8 to 8).

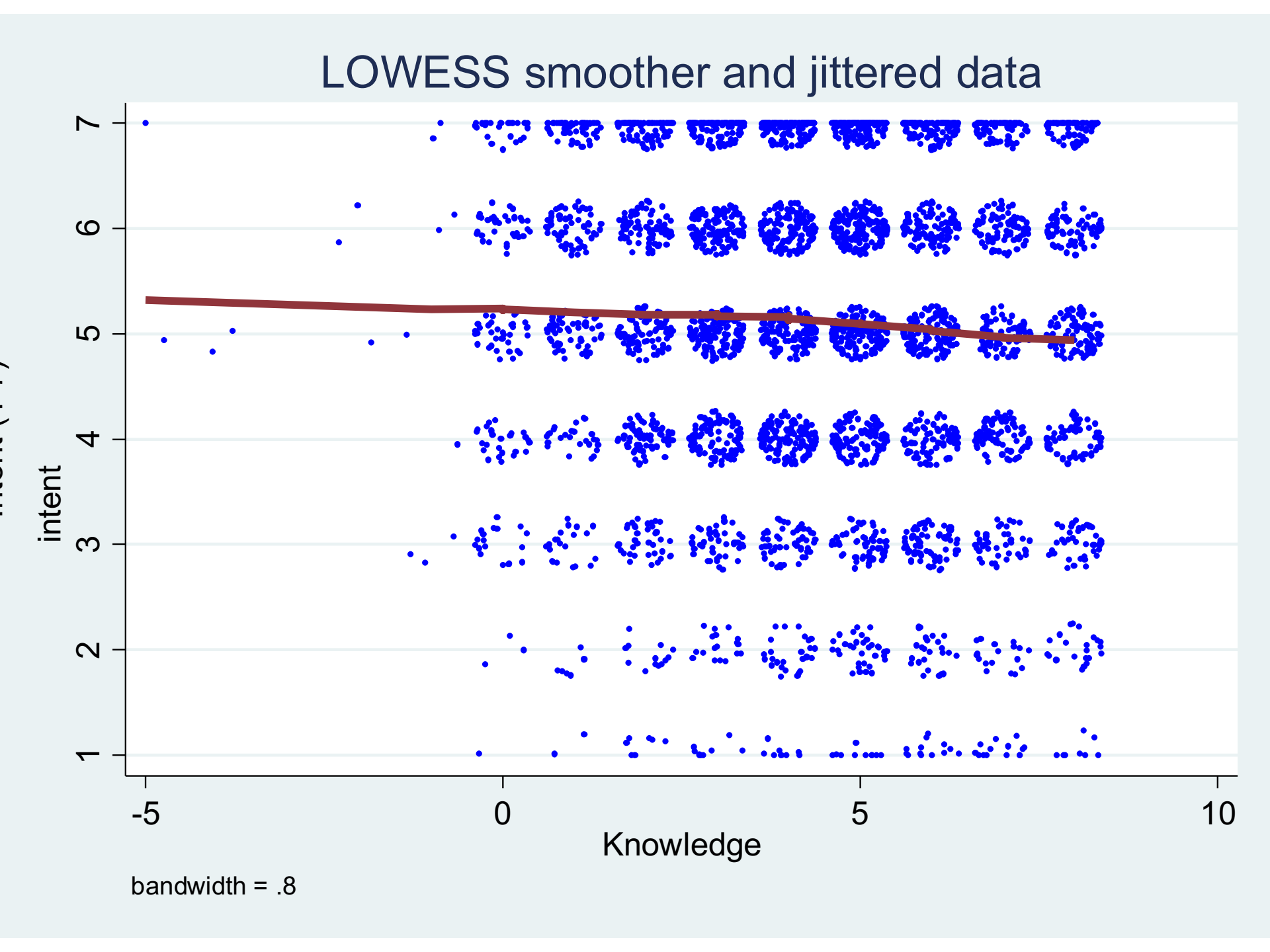
What type graph to assess association?

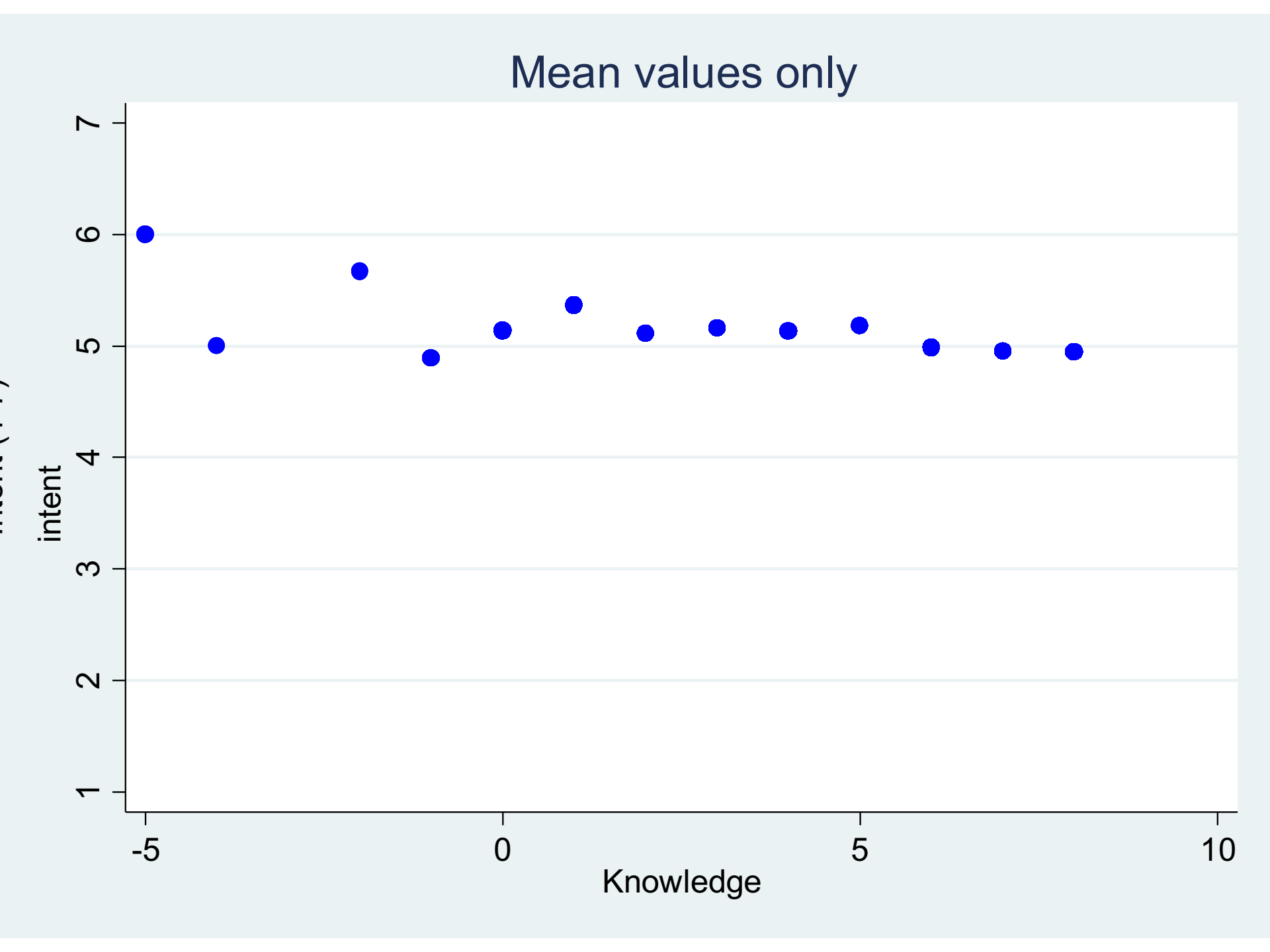
Data consistent with Pot

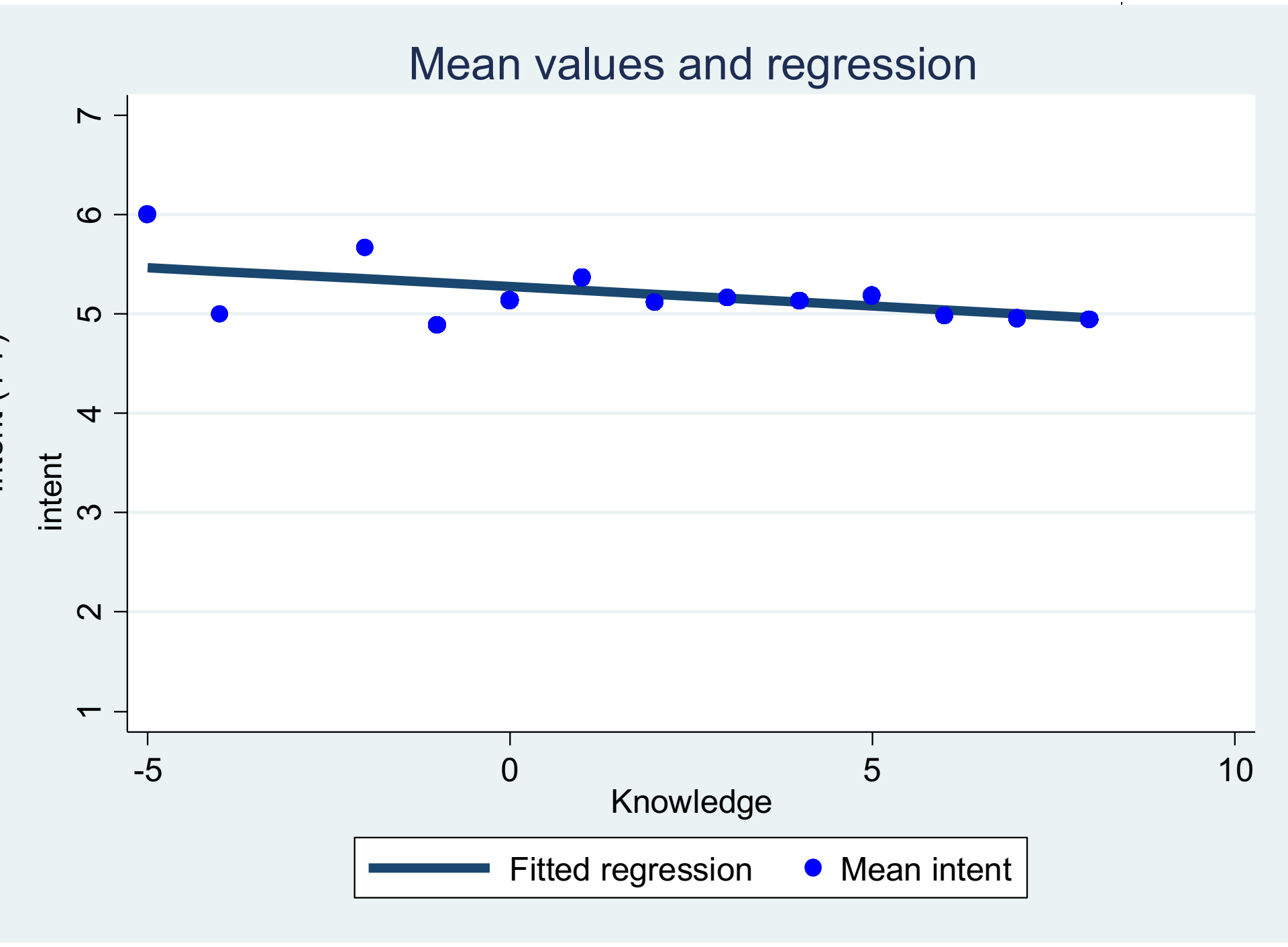




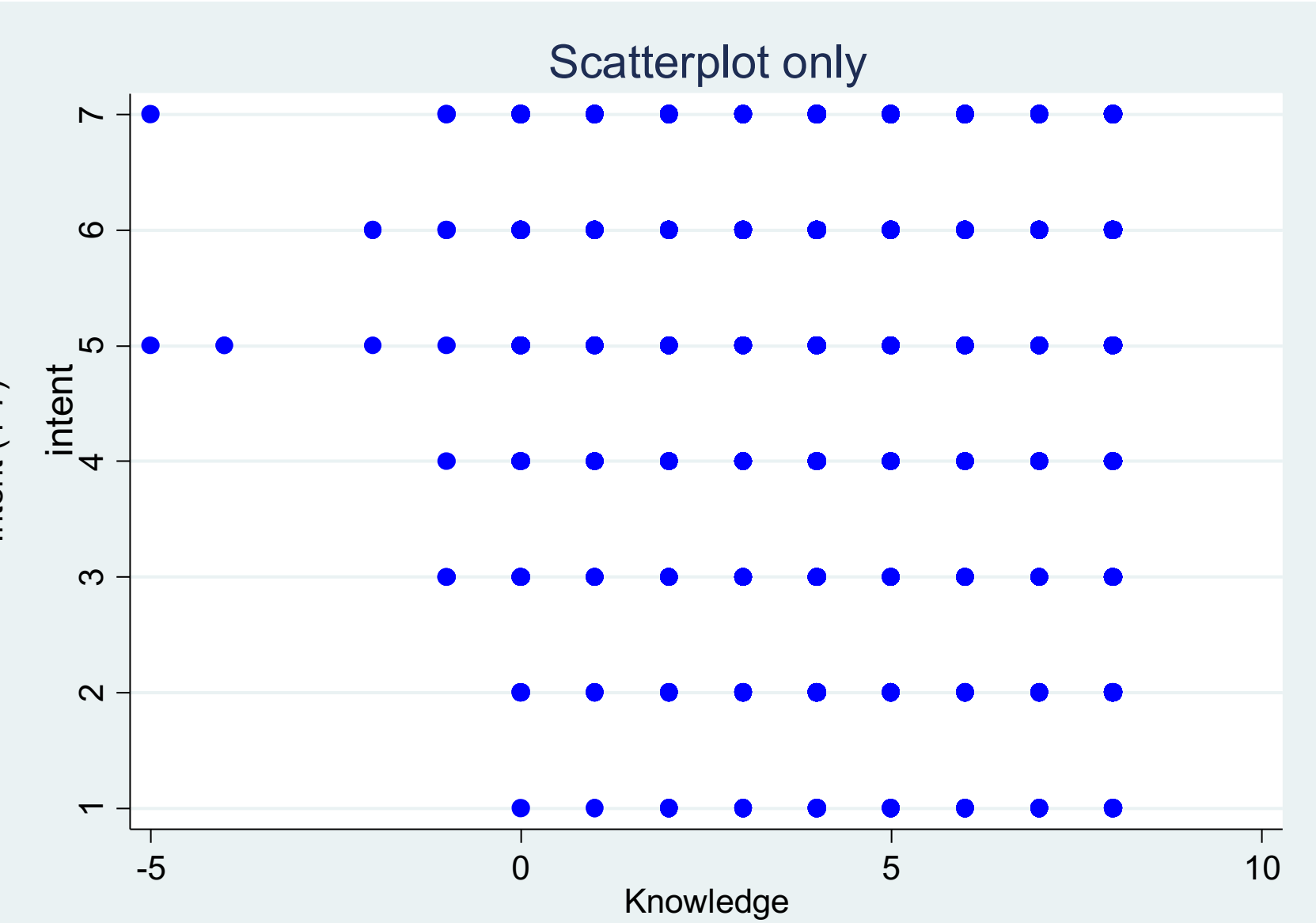

Scatterplot only
intent
0
1
2
3
4
5
6
7
8
-5
0
5
10
Knowledge

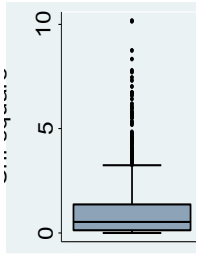

LOWESS smoother and jittered data
intent
1
2
3
4
5
6
7
-5
0
5
10
Knowledge
bandwidth = .8


Mean values only
intent
1
2
3
4
5
6
7
-5
0
5
10
Knowledge

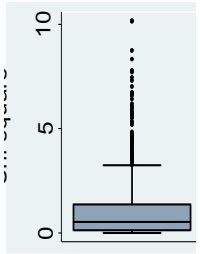

Mean values and regression
intent
1
2
3
4
5
6
7
-5
0
5
10
Knowledge
Fitted regression
Mean intent

Compare with original



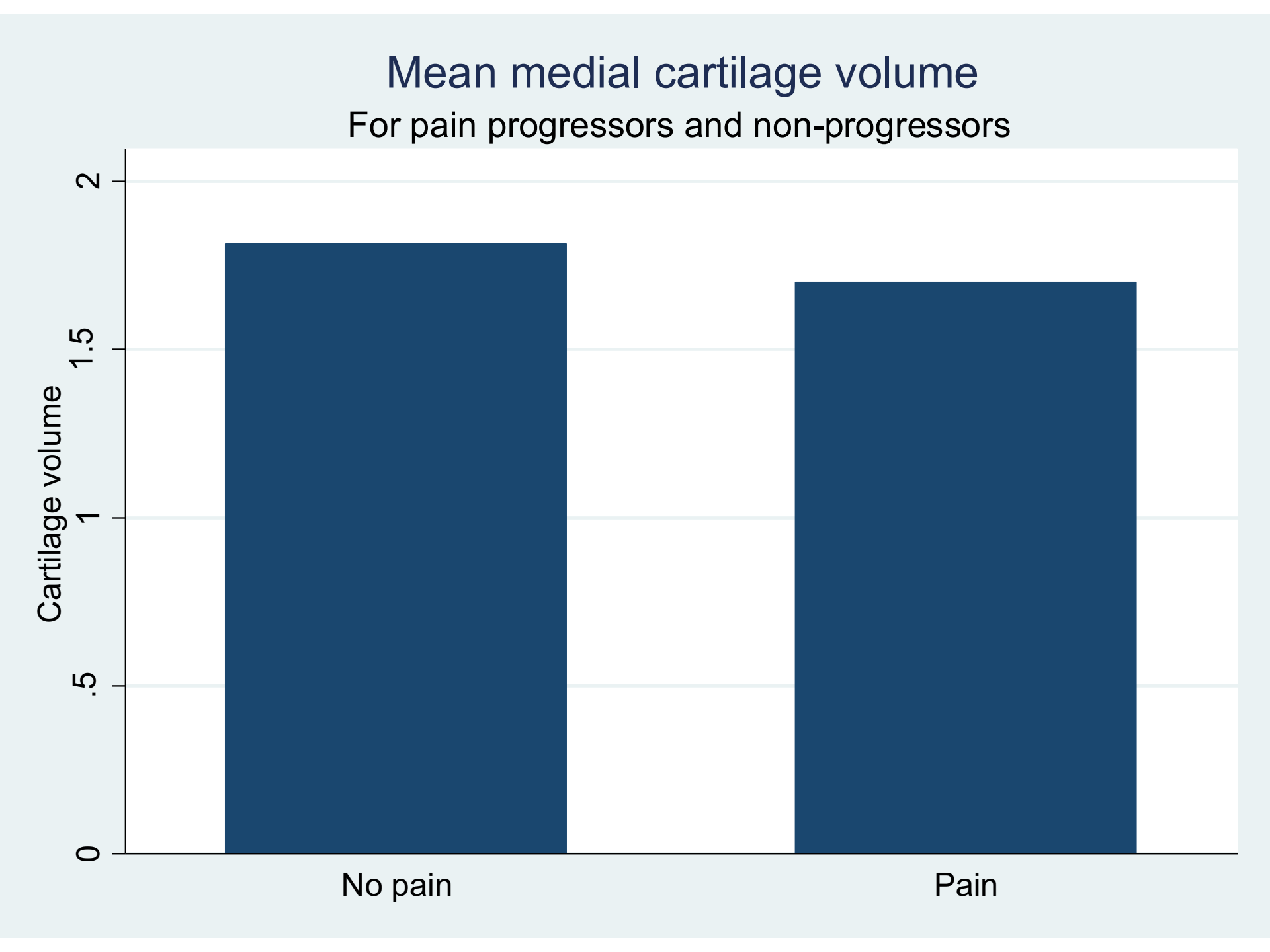


Graph practice



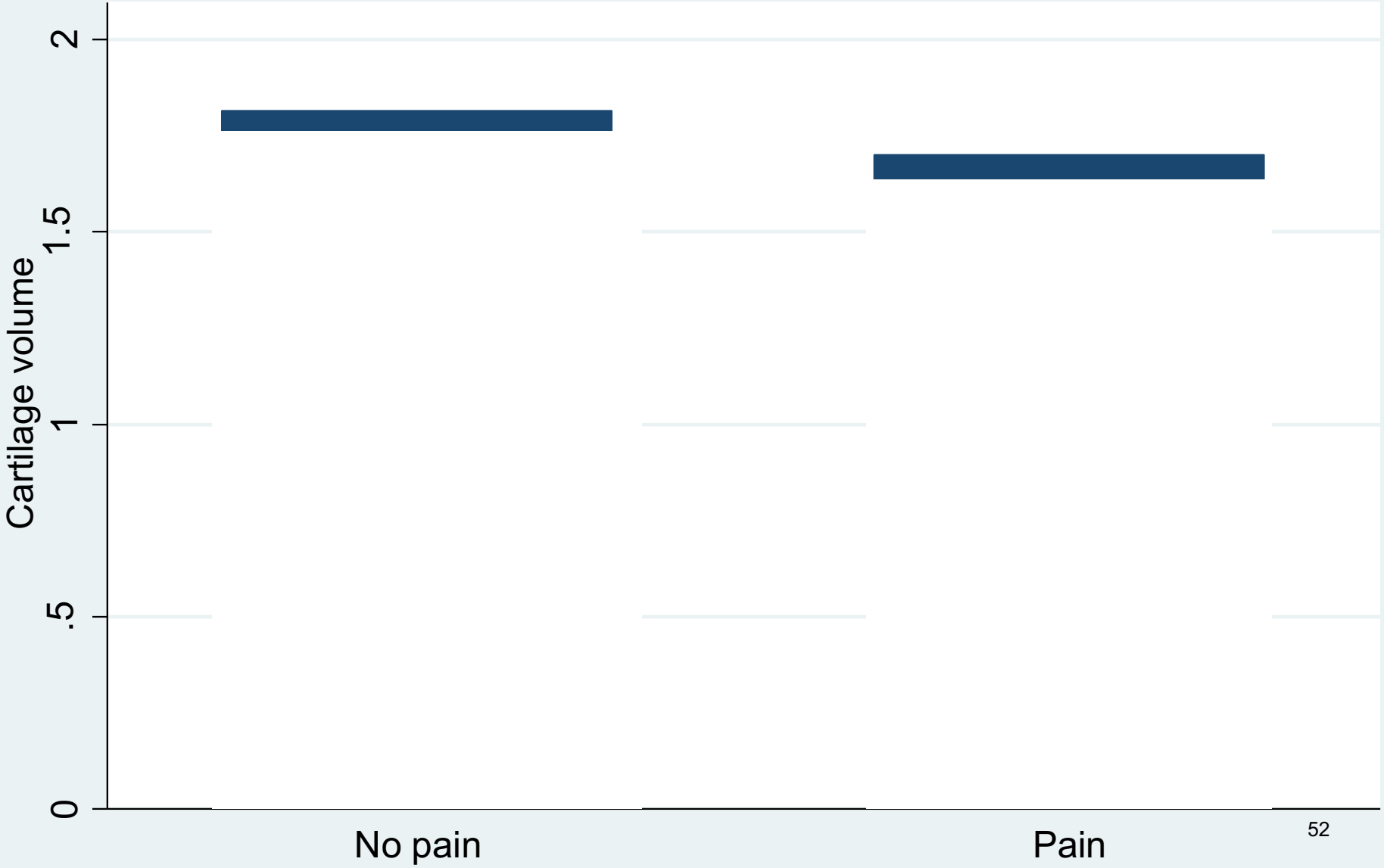
OAI data – is the volume of cartilage in the knee different between those who have painful knee arthritis and those that do not?

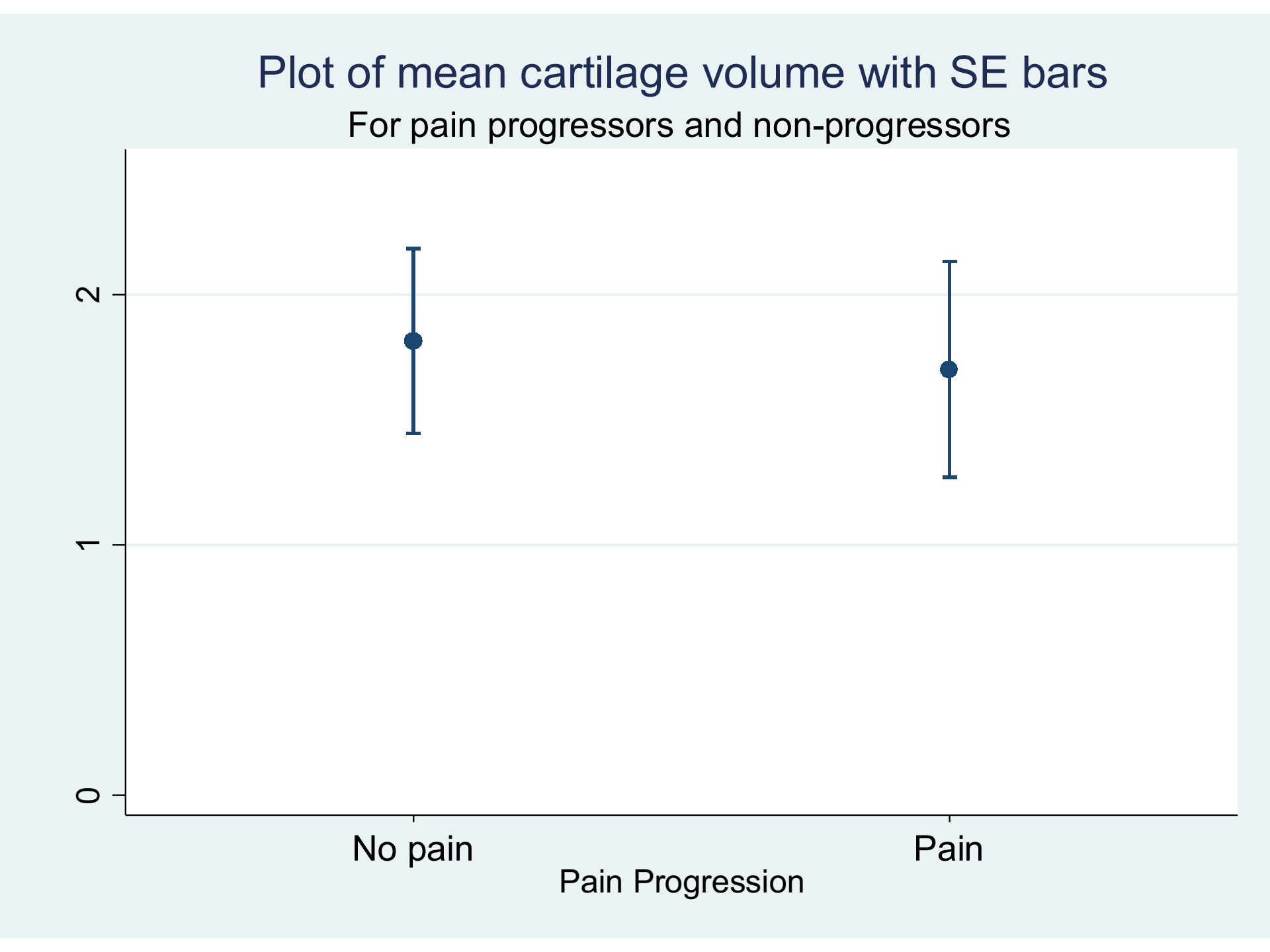
Preferred graph type?

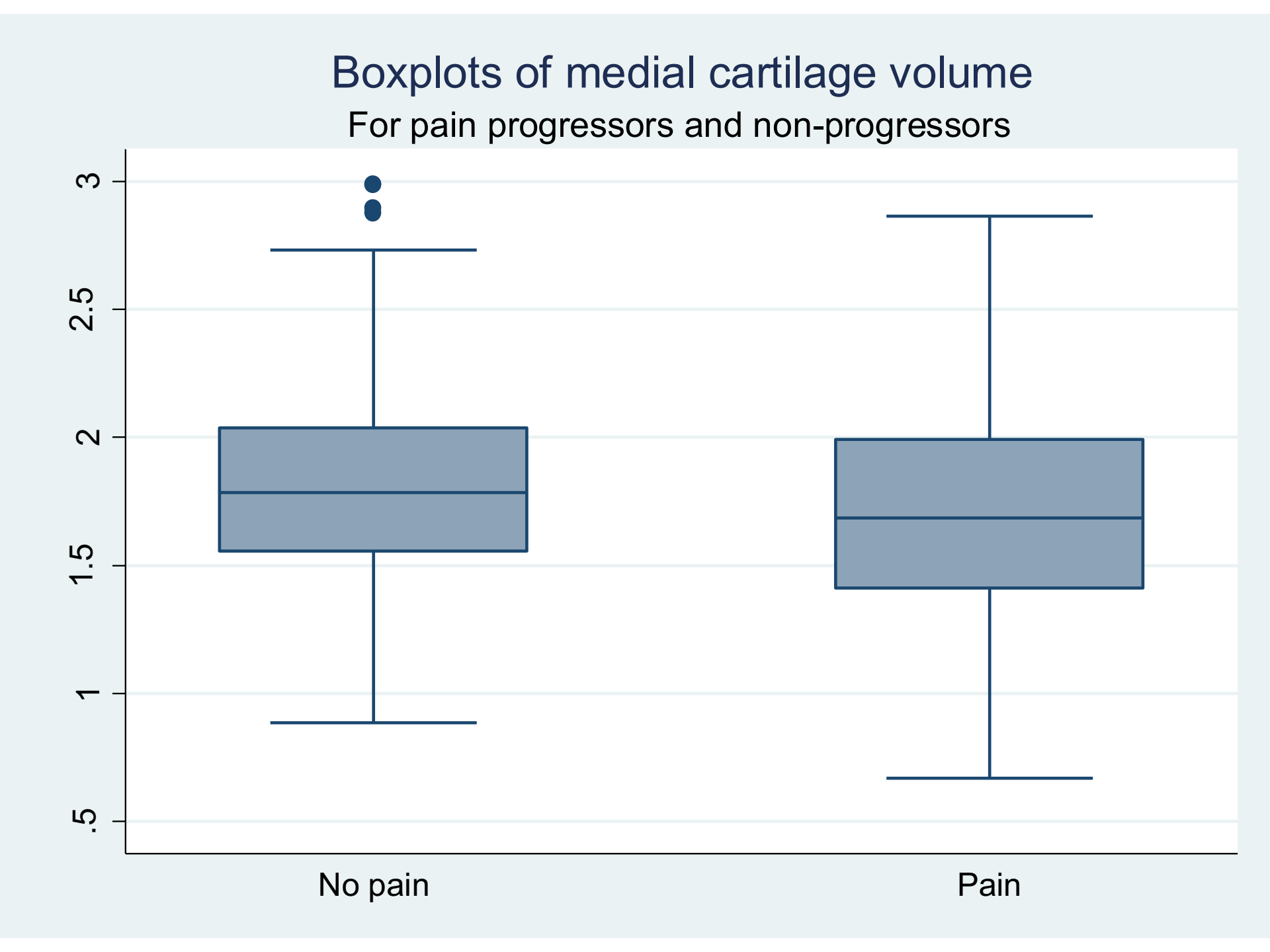

Mean medial cartilage volume
For pain progressors and non-progressors
Cartilage volume
0
.5
1
1.5
2
No pain
Pain

Mean medial cartilage volume

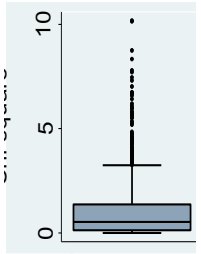
For pain progressors and non-progressors




Plot of mean cartilage volume with SE bars
For pain progressors and non-progressors
0
1
2
No pain
Pain
Pain Progression


Boxplots of medial cartilage volume
For pain progressors and non-progressors
3
2.5
2
1.5
1
.5
No pain
Pain

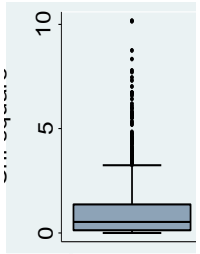
More variables at once (big data)



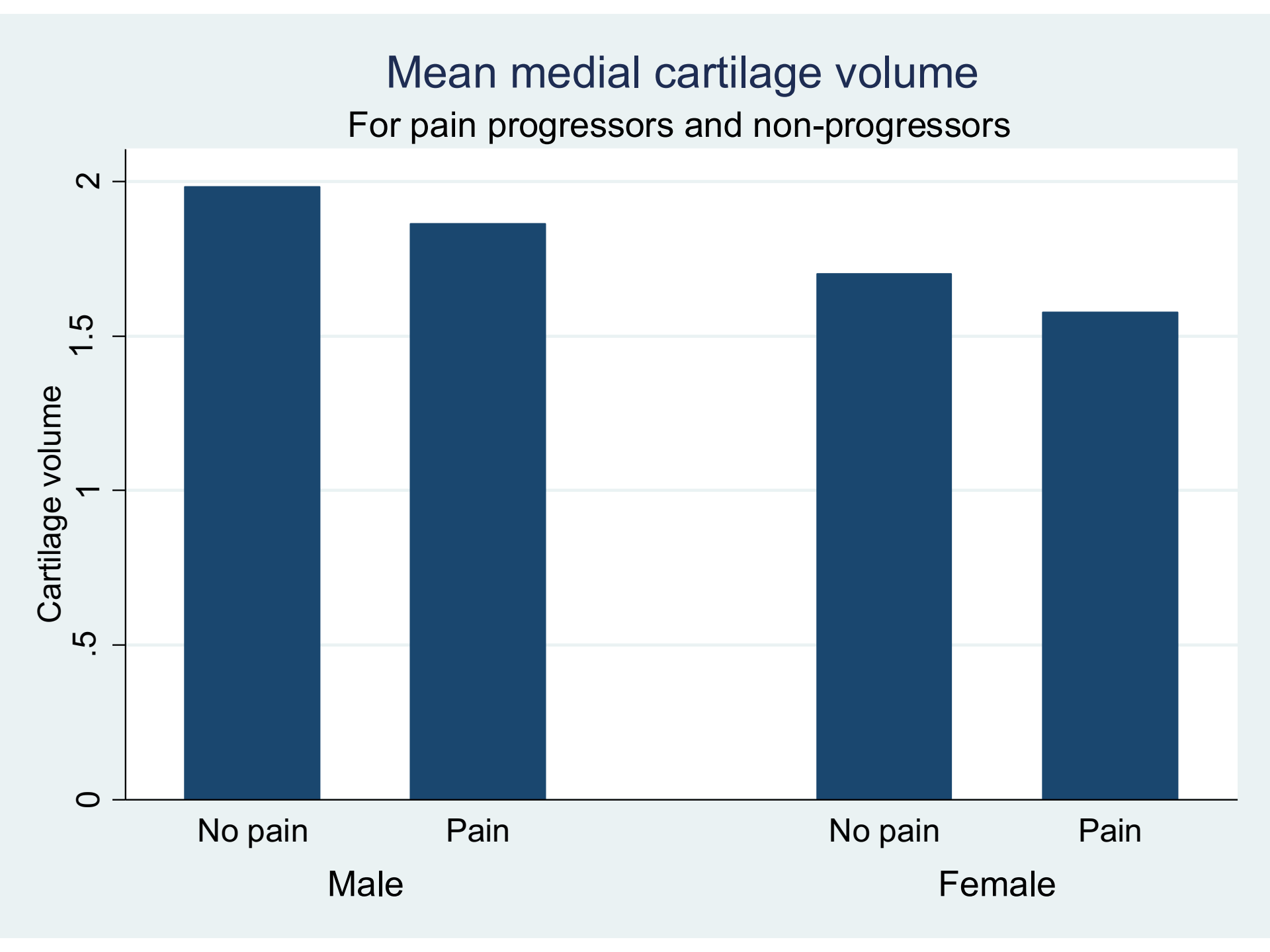
Adding additional variables to a two-variable graph

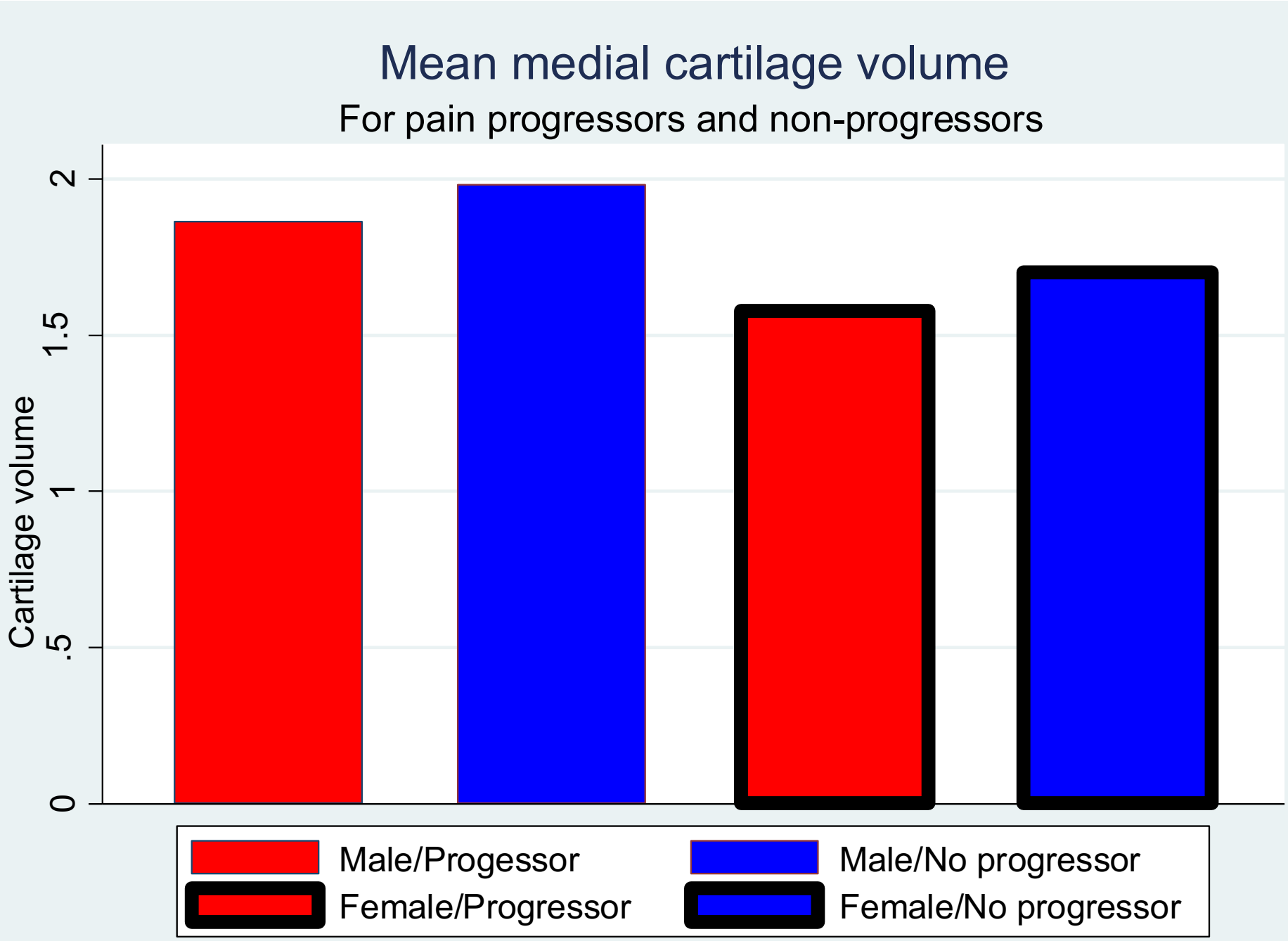
- Numeric
  - Symbol size
  - Symbol color shading
- Categorical
  - Color
  - Symbol type or line type
  - Different panels

Graph practice

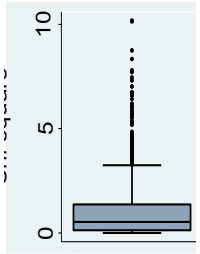


OAI data – is the volume of cartilage in the knee different between those who have painful knee arthritis and those that do not? And different between males and females?


Mean medial cartilage volume
For pain progressors and non-progressors
Cartilage volume
0
.5
1
1.5
2
No pain
Pain
No pain
Pain
Male
Female


Mean medial cartilage volume
For pain progressors and non-progressors
Cartilage volume
0
.5
1
1.5
2
Male/Progessor
Male/No progressor
Female/Progressor
Female/No progressor

Graph practice



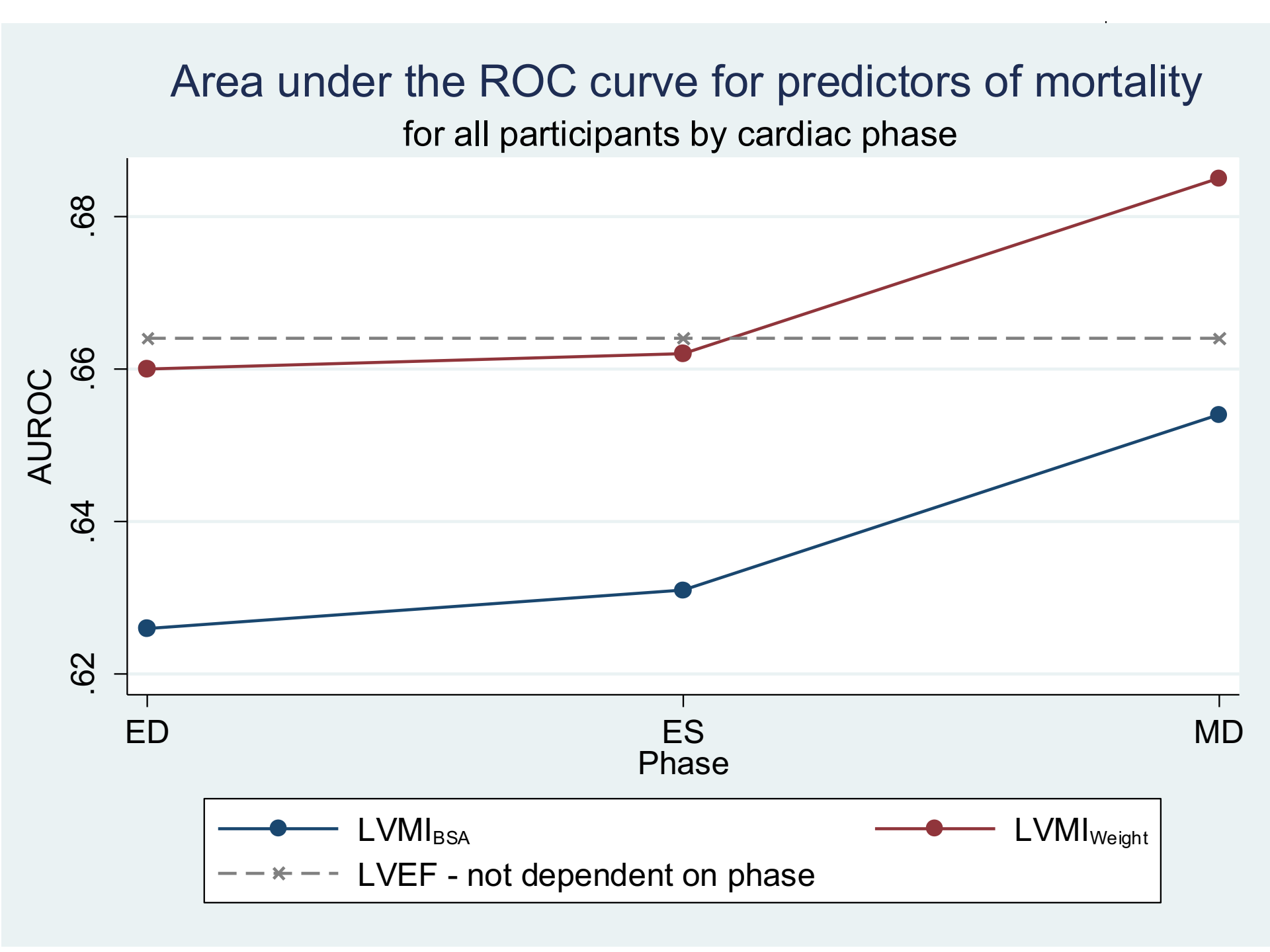
Back to the AUROC predictor data. Can we do better with a graph?

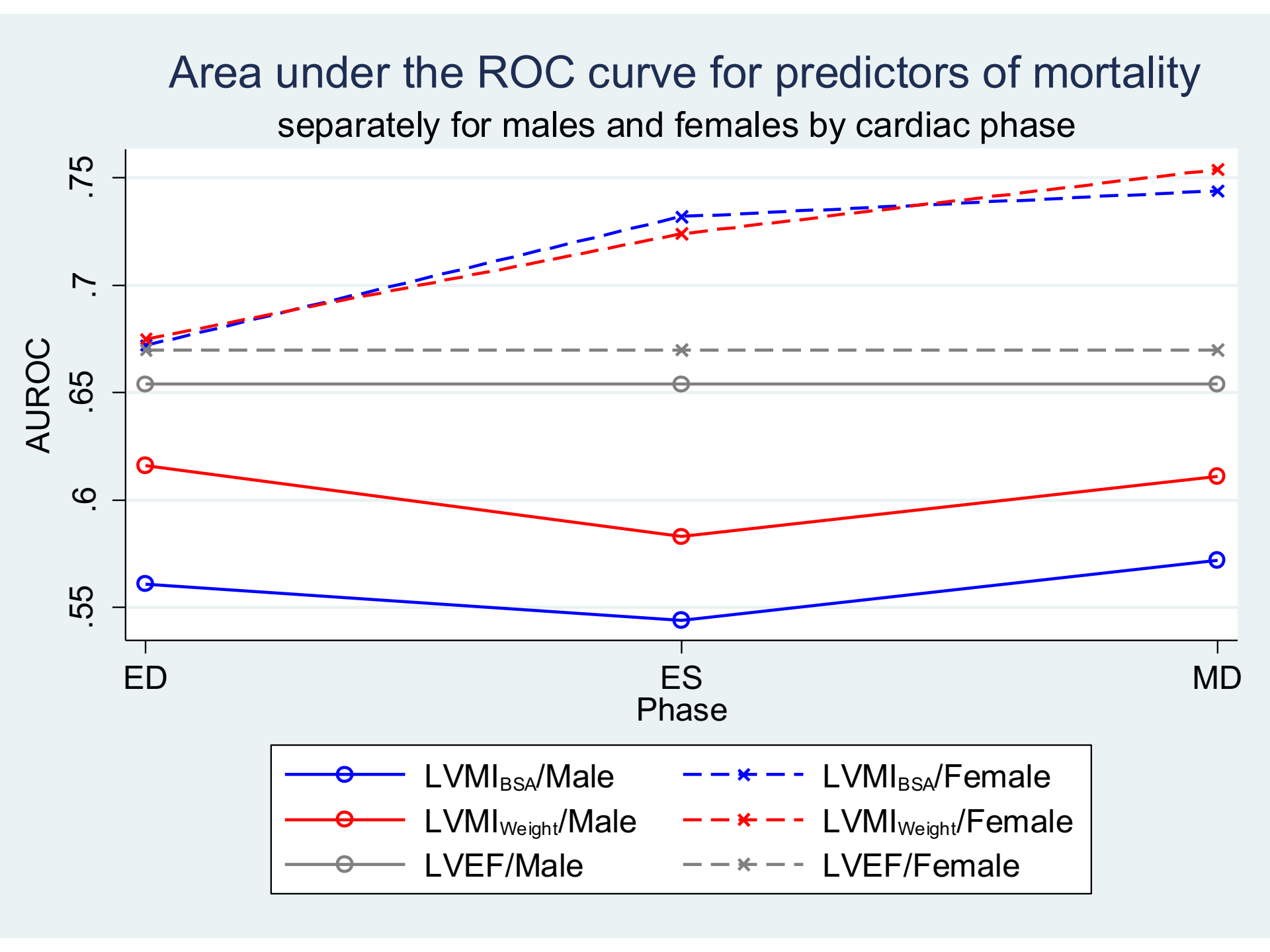
Preferred graph type to display predictor, phase and AUROC?

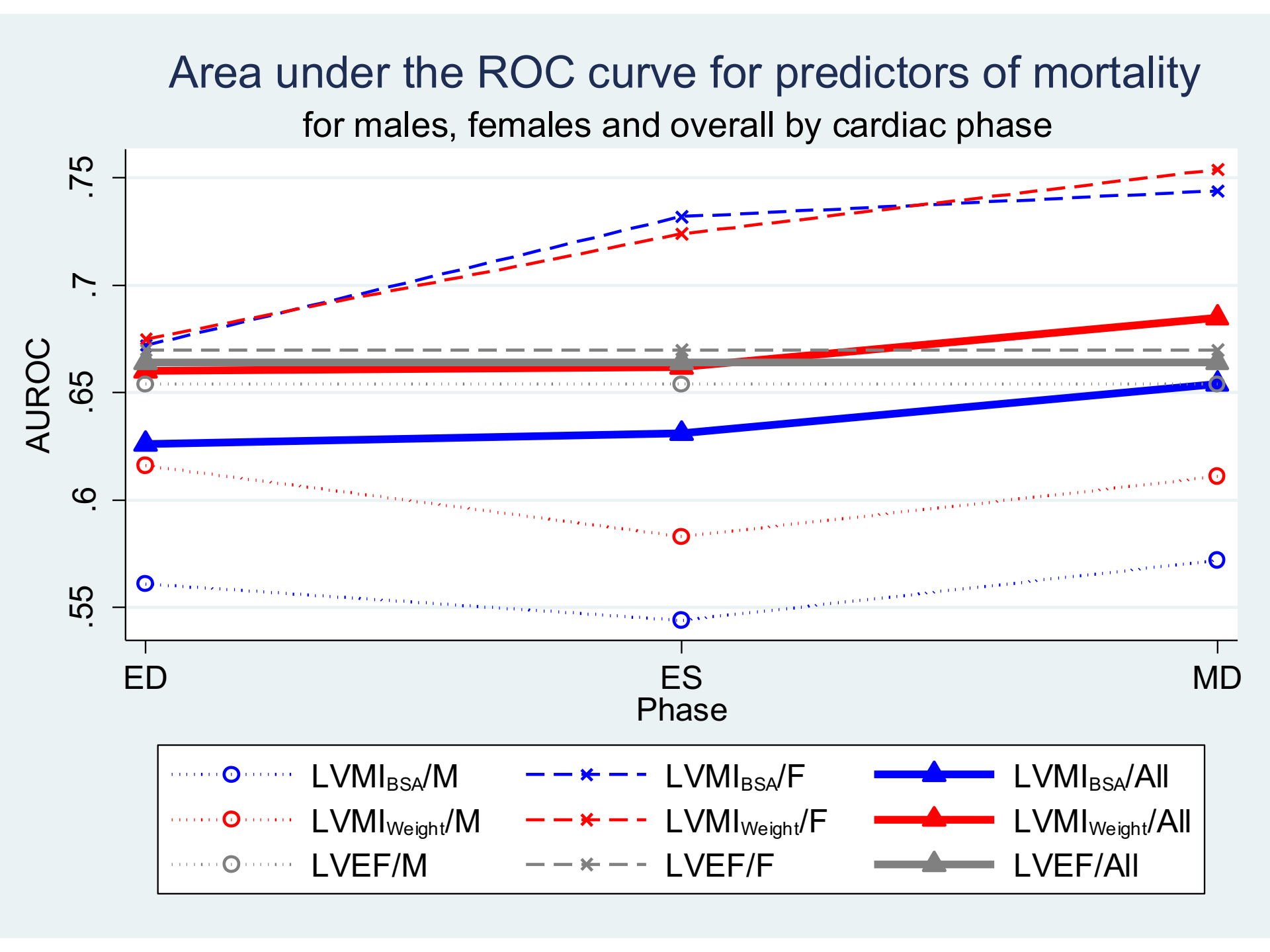
Predictor: categorical

Phase: categorical

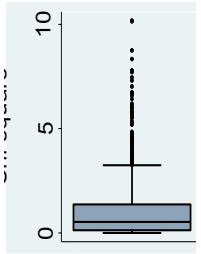
AUROC: numerical


Area under the ROC curve for predictors of mortality
for all participants by cardiac phase
AUROC
.62
.64
.66
.68
ED
ES
MD
Phase
LVMI$_{BSA}$
LVMI$_{Weight}$
LVEF - not dependent on phase


Area under the ROC curve for predictors of mortality
separately for males and females by cardiac phase
AUROC
.55
.6
.65
.7
.75
ED
ES
MD
Phase
LVMI$_{BSA}$/Male
LVMI$_{BSA}$/Female
LVMI$_{Weight}$/Male
LVMI$_{Weight}$/Female
LVEF/Male
LVEF/Female

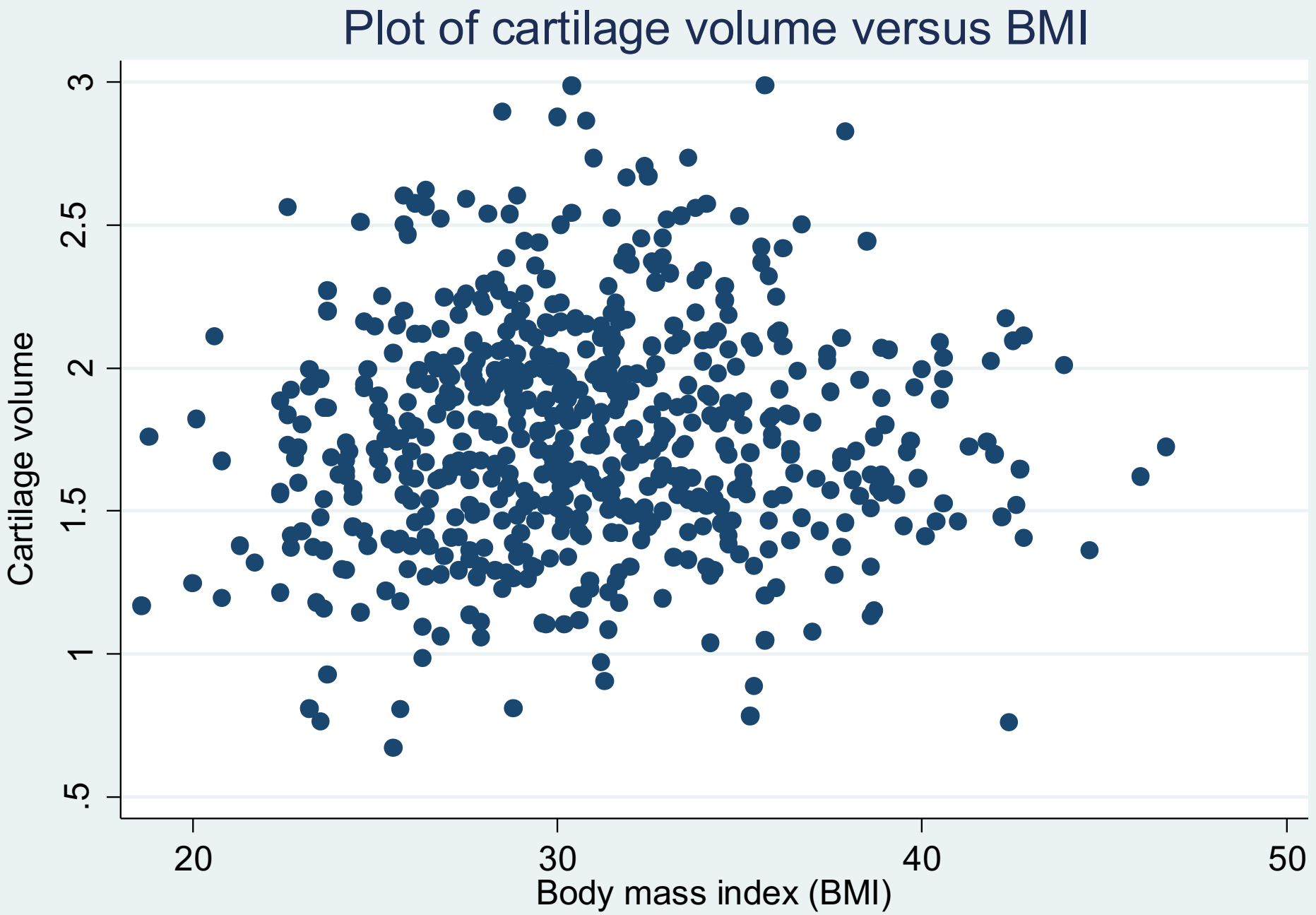

Area under the ROC curve for predictors of mortality
for males, females and overall by cardiac phase
AUROC
.55
.6
.65
.7
.75
ED
ES
MD
Phase
LVMI$_{BSA}$/M
LVMI$_{BSA}$/F
LVMI$_{BSA}$/All
LVMI$_{Weight}$/M
LVMI$_{Weight}$/F
LVMI$_{Weight}$/All
LVEF/M
LVEF/F
LVEF/All

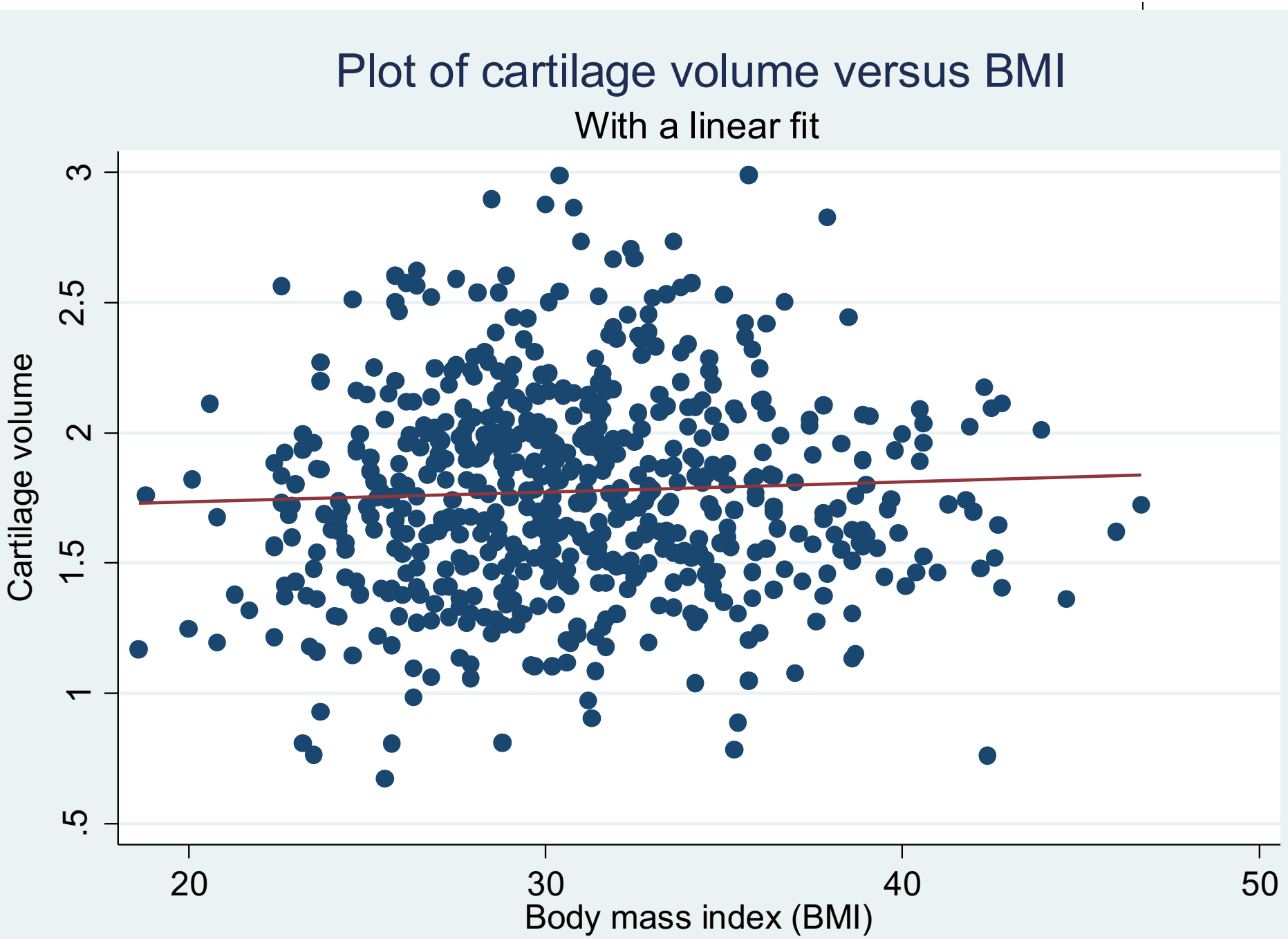
Graph practice

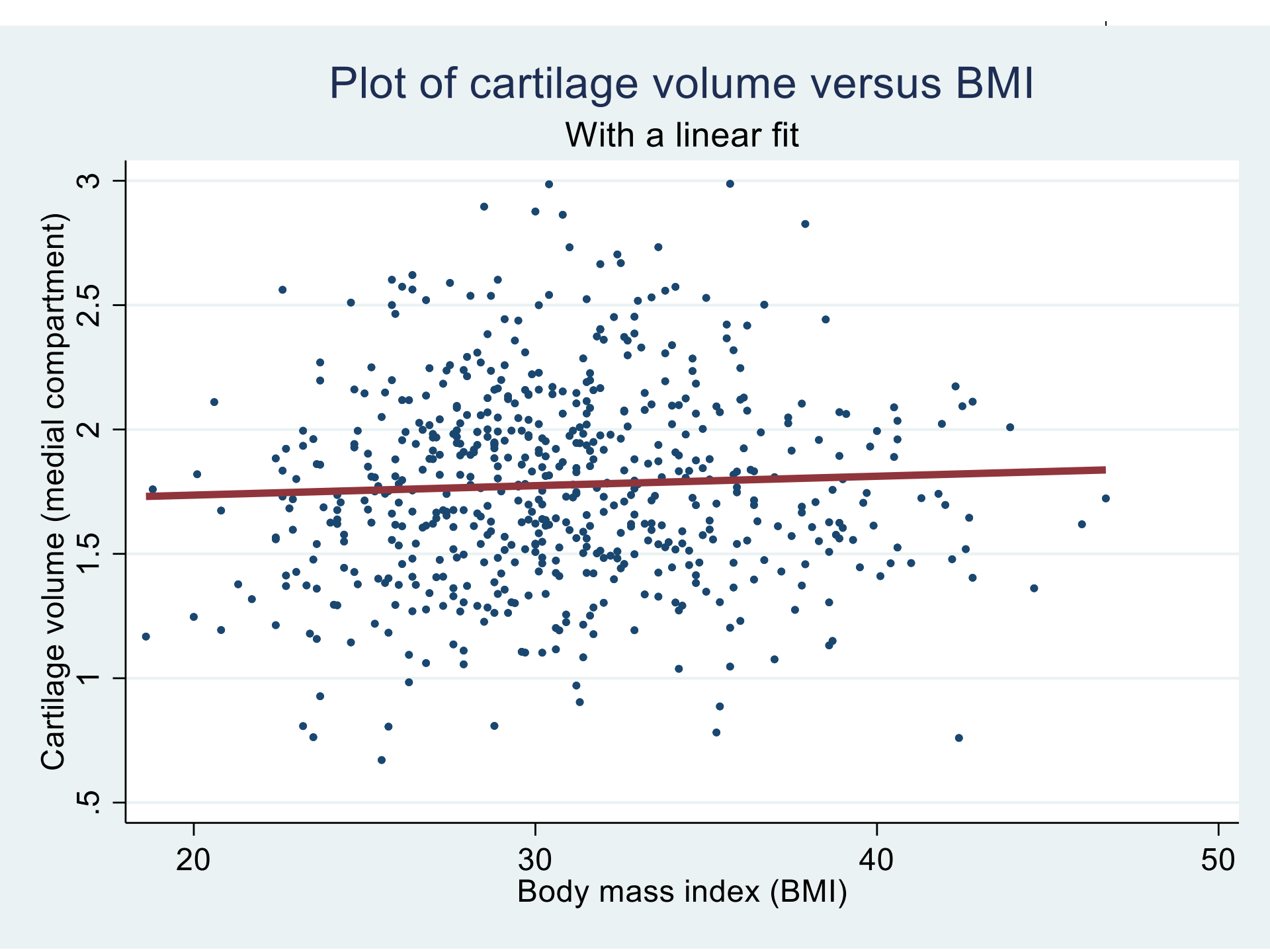


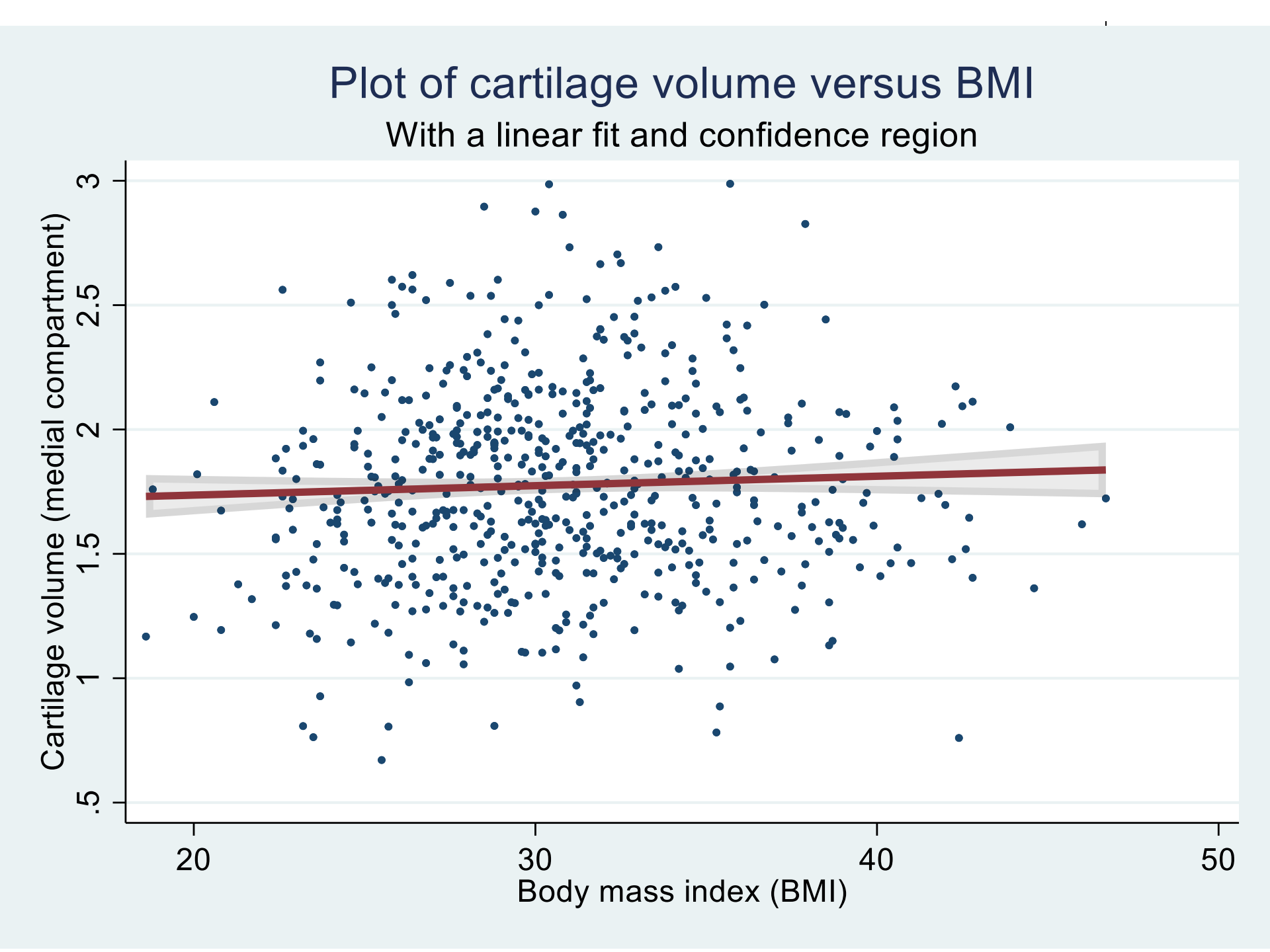
OAI data – is the volume of cartilage in the knee related to body mass index (BMI)?

Preferred graph type?

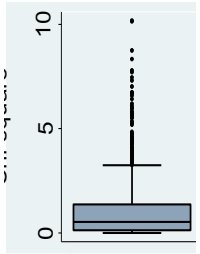

Plot of cartilage volume versus BMI
Cartilage volume
.5
1
1.5
2
2.5
3
20
30
40
50
Body mass index (BMI)


Plot of cartilage volume versus BMI
With a linear fit
Cartilage volume
.5
1
1.5
2
2.5
3
20
30
40
50
Body mass index (BMI)


Plot of cartilage volume versus BMI
With a linear fit
Cartilage volume (medial compartment)
.5
1
1.5
2
2.5
3
20
30
40
50
Body mass index (BMI)


Plot of cartilage volume versus BMI
With a linear fit and confidence region
Cartilage volume (medial compartment)
.5
1
1.5
2
2.5
3
20
30
40
50
Body mass index (BMI)

Graph practice

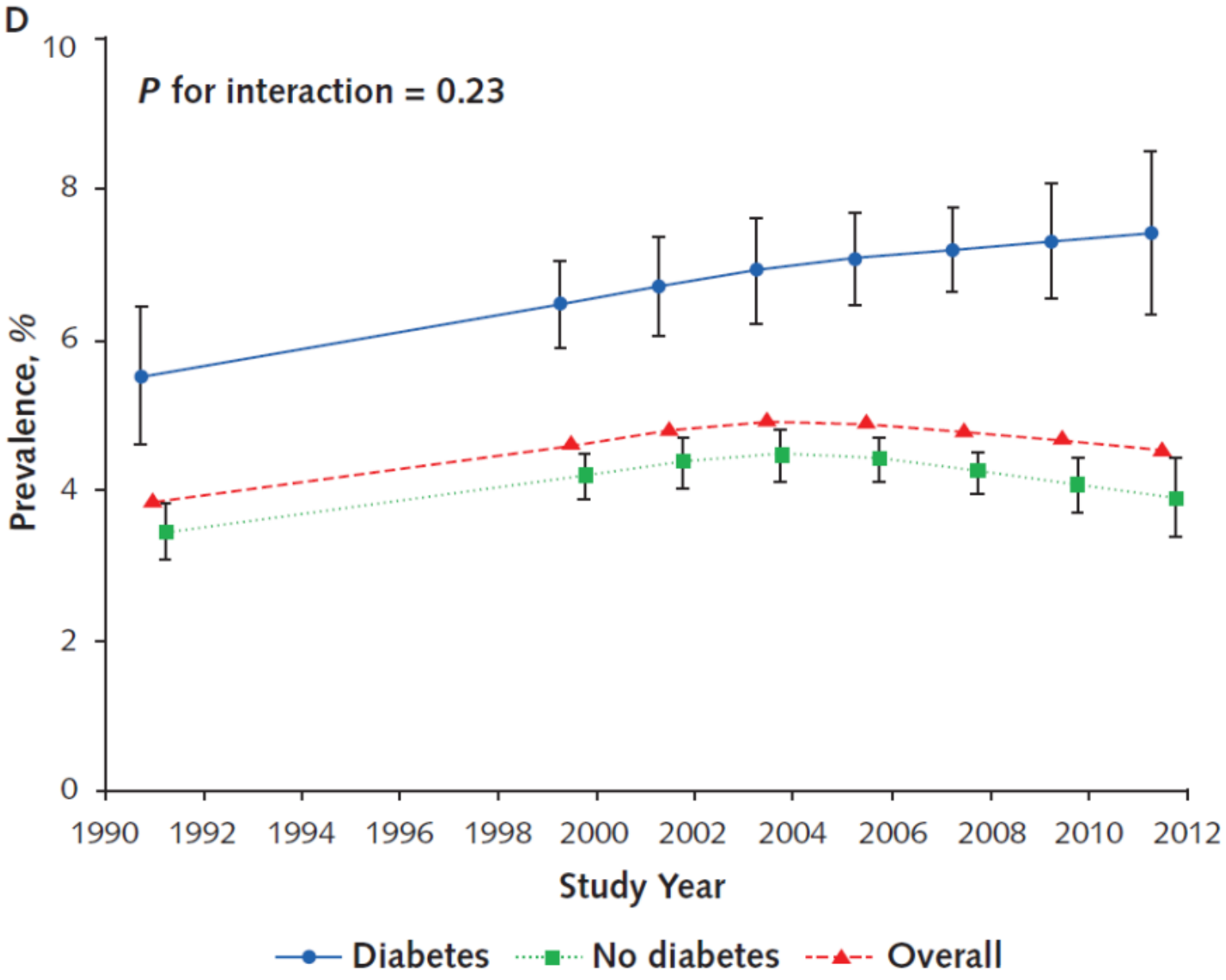


NHANES data – is **trend** in chronic kidney disease prevalence over **time**?

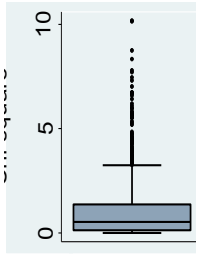
Does it differ between diabetics and non-diabetics?

Preferred graph type? Hint… trend.

Figure. Adjusted prevalence of stage 3 and 4 chronic kidney disease (estimated glomerular filtration rate of 15 to 59 mL/min/1.73 m^2 calculated with Chronic Kidney Disease Epidemiology Collaboration equation) in U.S. adults, by age (*A*), sex (*B*), race/ethnicity (*C*), and presence or absence of diabetes (*D*), NHANES 1988–1994 through 2011–2012.

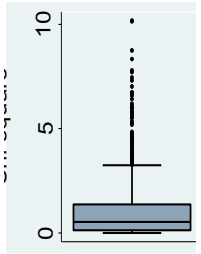


Interactive plots



- Very cool! Very labor intensive
- Consider if this is worth your effort
- If you need to reach a large audience – your work is important and you should budget resources to assist you here

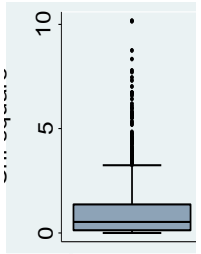
Other kinds of graphics



- Chronic Kidney disease map and animation

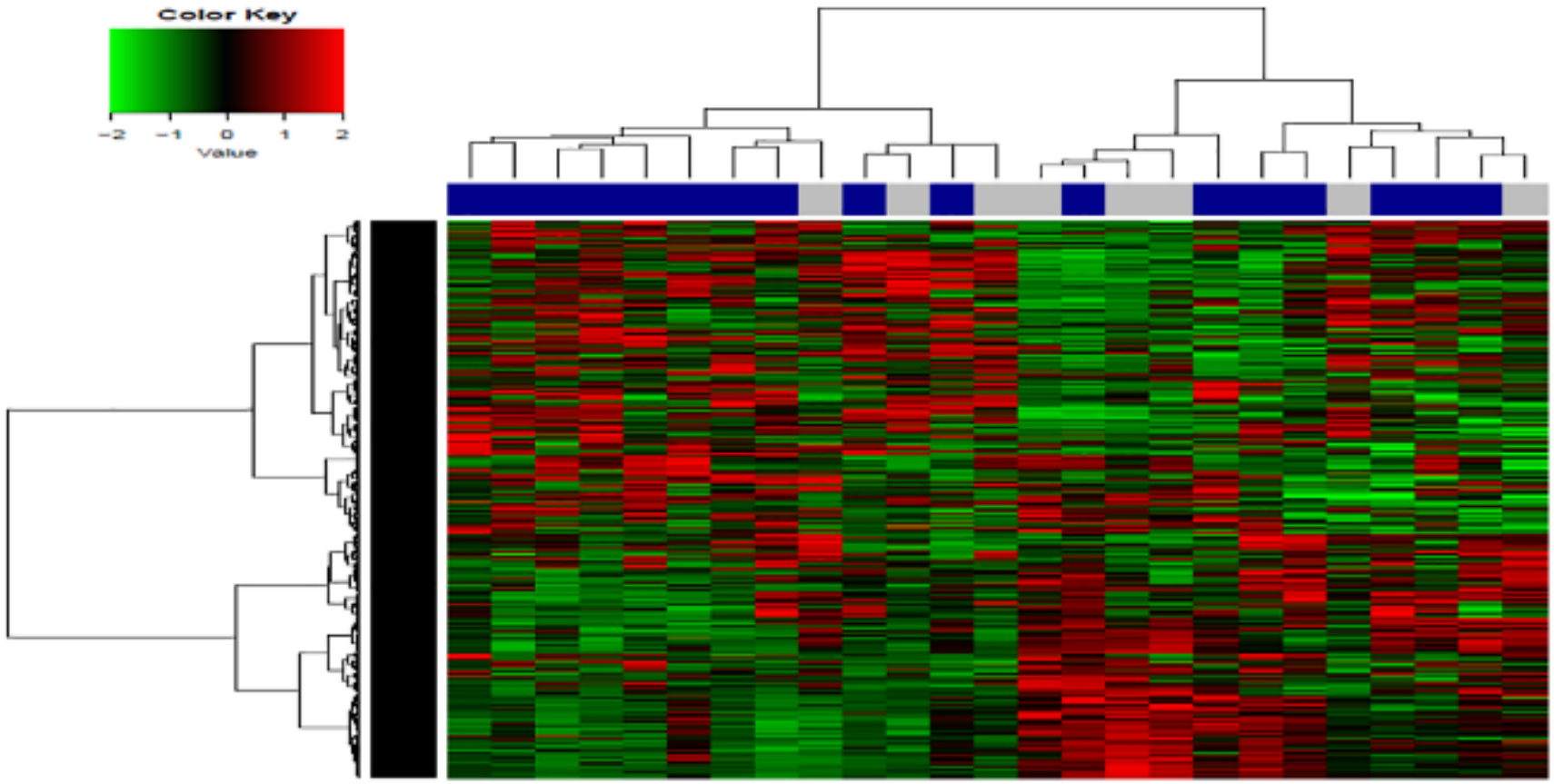
https://nccd.cdc.gov/CKD/Detail.aspx?QNum=Q69&Year=2016&colors=all#refreshPosition

Other kinds of graphics



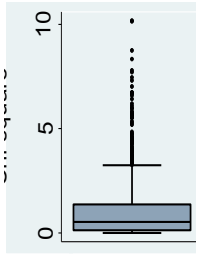
- Genetic heatmaps

Fig 2. NOJAH genome-wide heatmap (GWH) analysis of RNA-Seq derived gene expression data using TCGA BRCA expression dataset.



Rupji M, Dwivedi B, Kowalski J (2019) NOJAH: NOt Just Another Heatmap for genome-wide cluster analysis. PLOS ONE 14(3): e0204542. https://doi.org/10.1371/journal.pone.0204542
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204542

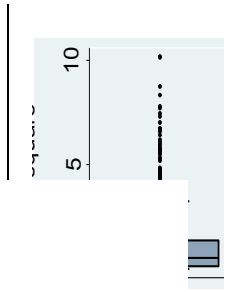
Other kinds of graphics



- McCulloch co-authors on UCSF Profiles

https://profiles.ucsf.edu/display/176718/network/coauthors/cluster

Summary



- Use tables for specific values, quantitative comparisons, precision, individual+summary information, differing scales.
- Use graphs to convey information quickly, for shapes, patterns and trends.
- Both should be self-explanatory and use a light touch on non-data
- Tables: compare columns, round aggressively, order purposefully
- Graphs: maximize data to ink, use color, symbols and line types purposefully.