



---

# Family Medicine

## SPSS: Part two

Melissa Pirrie

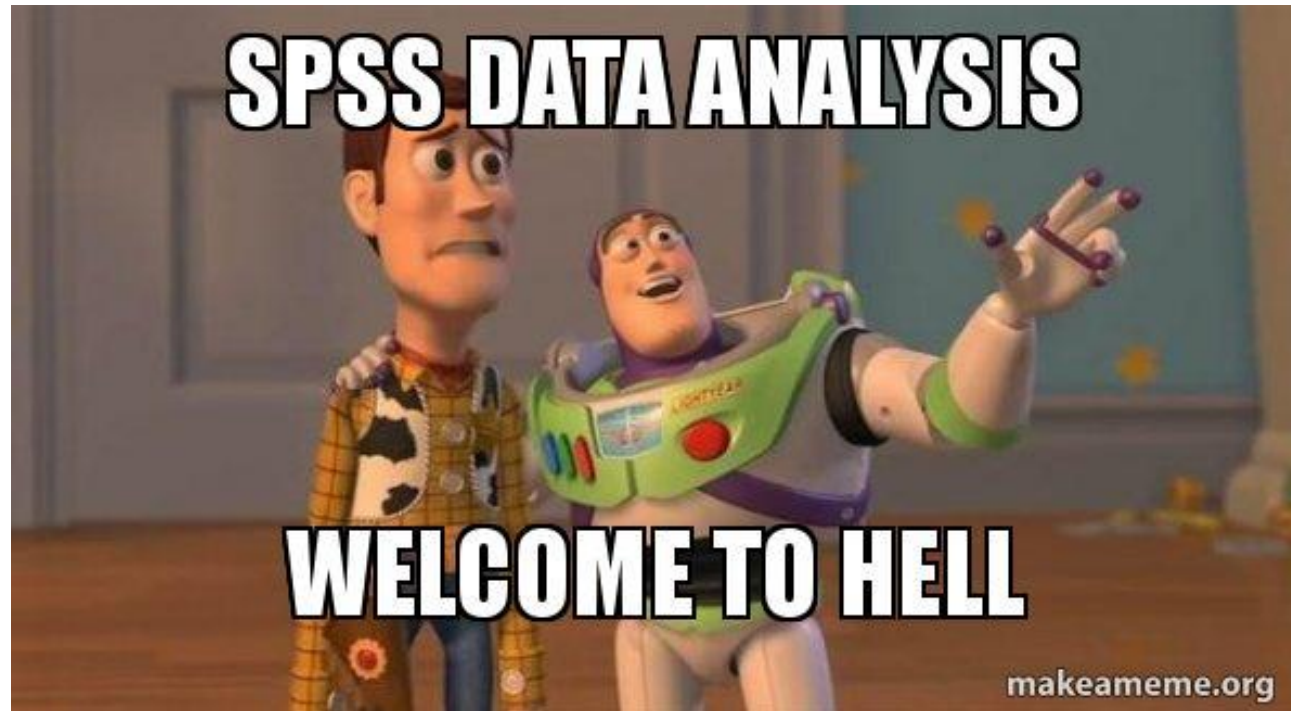
Larkin Lamarche

@PresentersTwitterHandles

@McMasterFamMed

# Learning objectives

- Apply knowledge about checking database and data to SPSS output



# I spy: The research edition

- Study description
- Output from data cleaning
- Spot the possible issues
  - Is this something systematic or random?
  - Why do you think it happened (think about the two views in SPSS: 'Variable view' and 'Data view')
  - How would you correct it?



# I spy: Scenario 1

**Statistics**

		Control or Intervention?	Site	How would you describe your ethnicity?
N	Valid	50	51	29
	Missing	1	0	22

Random (points to Site=51)  
Systematic (points to Site=0)

## Frequency Table

**Control or Intervention?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1	2.0	2.0	2.0
	Intervention	23	45.1	46.0	48.0
	Control	26	51.0	52.0	100.0
	Total	50	98.0	100.0	
Missing	System	1	2.0		
Total		51	100.0		

← (points to Valid=0, Frequency=1)

- Random
- Data entry error
- Find and correct in 'data view'

**Site**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Winter place	29	56.9	56.9	56.9
	Summer place	18	35.3	35.3	92.2
	The good place	2	3.9	3.9	96.1
	999	2	3.9	3.9	100.0
	Total	51	100.0	100.0	

← (points to Valid=999, Frequency=2)

- Systematic
- Missing values not specified
- Correct in 'variable view'

# I spy: Scenario 1

How would you describe your ethnicity?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	European/White	23	45.1	79.3	79.3
	Other	5	9.8	17.2	96.6
	Don't Know	1	2.0	3.4	100.0
	Total	29	56.9	100.0	
Missing	System	22	43.1		
Total		51	100.0		

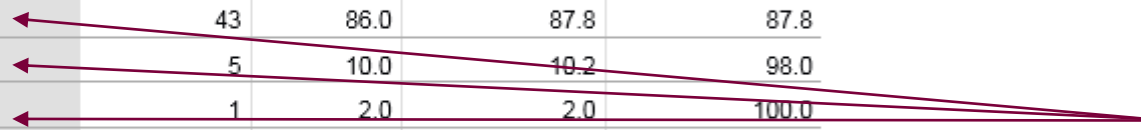
- Systematic, need to look into other patterns that might explain
  - Are they all from one site?
  - Was the question on the back of the page and missed by many?
  - Was the variable already recoded from a larger list of options, there was an error in the recoding?



What is your preferred language?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	43	86.0	87.8	87.8
	2	5	10.0	10.2	98.0
	3	1	2.0	2.0	100.0
	Total	49	98.0	100.0	
Missing	System	1	2.0		
Total		51	100.0		

- Systematic
- No value labels
- Fix in the 'variable view'



# I spy: Scenario 2

## Descriptives

	Descriptive Statistics				
	N	Minimum	Maximum	Mean	Std. Deviation
What is your age?	50	63	102	76.76	5.659
BMI	50	0	42	28.21	4.03
TUG_pre	50	6.1	99	33.41	8.58
P2	35	6.5	17.3	11.84	3.47
Satisfaction1_pre	50	1	10	6.20	2.05
Satisfaction2_pre	50	1	10	4.40	1.85
Satisfaction3_pre	50	1	10	4.85	1.95
Satisfaction_score_pre	50	1	10	5.20	2.15
Satisfaction1_post	51	1	10	6.30	2.07
Satisfaction2_post	51	1	10	7.14	1.72
Satisfaction3_post	51	1	10	6.85	1.45
Satisfaction_score_post	51	1	30	21.12	7.25
Qualityoflife_score_T0	50	0	1.00	.8508	.12284
Qualityoflife_score_T6	50	0	1.00	.8575	.11940
Valid N (listwise)	35				

- Could be random or systematic:
  - Data entry error? [Random]
  - Entered '0' whenever missing? [Systematic]
- Mean and SD seem within expected range
- Likely random error
- Find and fix in the 'data view'
  
- Could be random or systematic:
  - Data entry error? [Random]
  - Entered '99' whenever missing? [Systematic]
- Mean seems is much higher than expected and 99 is commonly used for missing
- Likely need to set 99 as missing value
- Fix in the 'variable view'

# I spy: Scenario 2

## Descriptives

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
What is your age?	50	63	102	76.76	5.659
BMI	50	0	42	28.21	4.03
TUG_pre	50	6.1	99	33.41	8.58
P2	35	6.5	17.3	11.84	3.47
Satisfaction1_pre	50	1	10	6.20	2.05
Satisfaction2_pre	50	1	10	4.40	1.85
Satisfaction3_pre	50	1	10	4.85	1.95
Satisfaction_score_pre	50	1	10	5.20	2.15
Satisfaction1_post	51	1	10	6.30	2.07
Satisfaction2_post	51	1	10	7.14	1.72
Satisfaction3_post	51	1	10	6.85	1.45
Satisfaction_score_post	51	1	30	21.12	7.25
Qualityoflife_score_T0	50	0	1.00	.8508	.12284
Qualityoflife_score_T6	50	0	1.00	.8575	.11940
Valid N (listwise)	35				

- Systematic
- No variable label
- Fix in 'variable view'

- Systematic  
(Post-score mean and SD suggest it isn't a random data entry of '30')
- Most likely: pre-score is a mean of the 3 satisfaction questions and post-score is a sum of the 3 satisfaction questions
- Re-compute the variable

# I spy: Scenario 3

Pre-Intervention:

Reliability Statistics	
Cronbach's Alpha	N of Items
.812	3

Post-Intervention:

Reliability Statistics	
Cronbach's Alpha	N of Items
.529	3

- Much lower than pre-intervention
- Reverse coding?

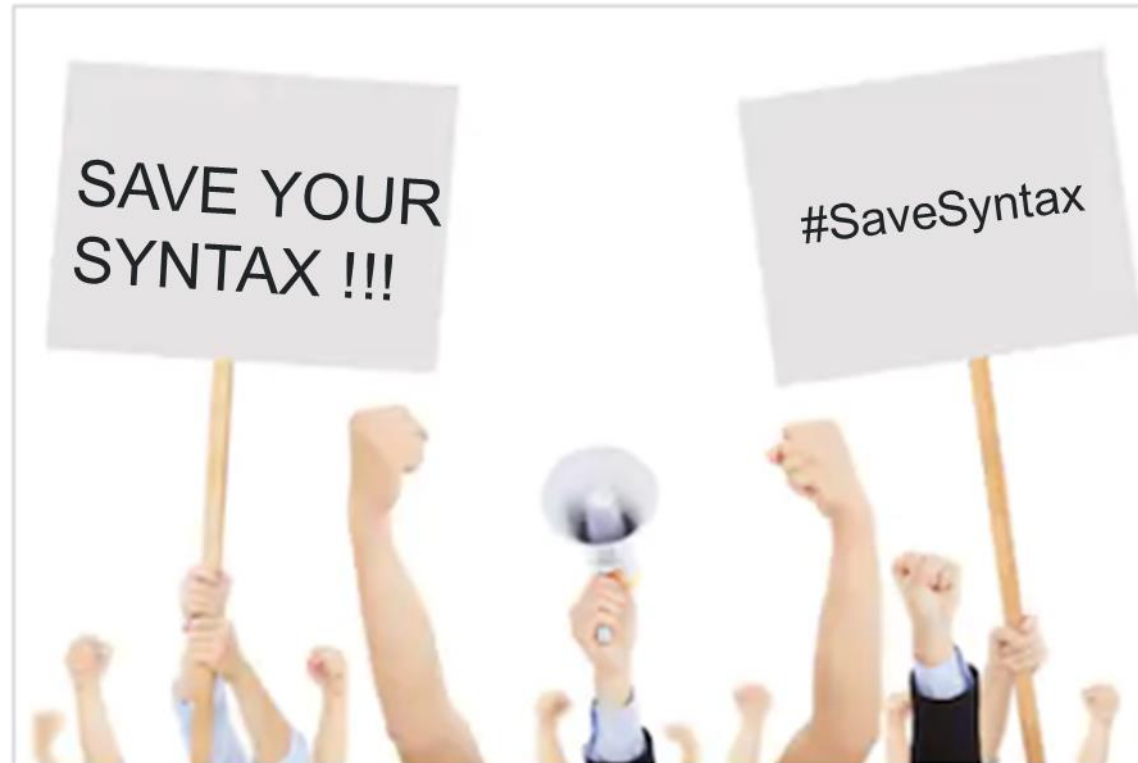




# I spy: Bonus

**BONUS:** You need to recode your continuous age variable into categories for future analyses. You use "Transform -> Recode into Different Variables" to tell SPSS how you want age to be recoded and get the syntax below in the output window; what should you do with it?

```
RECODE Age (55 thru 64=1) (65 thru 74=2) (75 thru 84=3) (85 thru Highest=4) INTO  
Age_Category.  
VARIABLE LABELS Age_Category 'Age Category'.
```



# Assumptions about SPSS

- You know your measures well
  - How are items measured
  - Scoring, calculations needed
  - Units
- You have a basic understanding of qualities of data
  - Variable types
  - Descriptive statistic meanings
    - Mean, standard deviation, range, etc
- You have basic understanding on statistical analyses
  - Which test you need to answer which type of research question
  - Which type of test can be run for which type of data



# Family Medicine

Department of Family Medicine  
Michael G. DeGroote School of Medicine  
Faculty of Health Sciences

[fammedmcmaster.ca](http://fammedmcmaster.ca)  
[@McMasterFamMed](https://twitter.com/McMasterFamMed)