

# Automated Data Summarization & Data Visualization for SPSS File in R

Idea & Process

# 2 Main Parts of SPSS File

## Data Structure (Variable View)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	Case_ID	Numeric	8	2	Case_ID	None	None	10
2	ID	String	3	0	ID.	None	None	6
3	Check_Gro...	Numeric	8	2	Check_Group	{1.00, Offli...	None	10
4	S2	Numeric	8	2	S2.ระบุมหา...	{18.00, 001...	None	10
5	S2Group	Numeric	8	2	S2Group.S2	{2.00, 18-24	None	10
6	Q11_1	Numeric	8	2	Q11_1.Time 1st	{1.00, 6.00 -	None	10
7	Q12_1	Numeric	8	2	Q12_1.Day 1st	{1.00, ...	None	10
8	Q13_1	Numeric	8	2	Q13. - 1	{1.00, ...	None	10
9	Q10Att1	Numeric	8	2	Q10. - ...	{1.00, ...	None	10
10	Q10Att2	Numeric	8	2	Q10. - ...	{1.00, ...	None	10
11	Q10Att3	Numeric	8	2	Q10. - ...	{1.00, ...	None	10
12	Q10Att4	Numeric	8	2	Q10. - ...	{1.00, ...	None	10
13	Q10Att5	Numeric	8	2	Q10. - ...	{1.00, ...	None	10

## Data Input (Data View)

	Case_ID	ID	Check_Group	S2	S2Group	Q11_1	Q12_1
1	2.00	2	4.00	24.00	2.00	5.00	7.00
2	3.00	3	1.00	37.00	4.00	.	.
3	6.00	6	3.00	42.00	4.00	5.00	1.00
4	7.00	7	3.00	27.00	3.00	5.00	1.00
5	8.00	8	1.00	18.00	2.00	.	.
6	9.00	9	3.00	27.00	3.00	5.00	1.00
7	10.00	10	3.00	22.00	2.00	5.00	7.00
8	11.00	11	3.00	18.00	2.00	6.00	8.00
9	12.00	12	3.00	18.00	2.00	5.00	1.00
10	13.00	13	3.00	24.00	2.00	5.00	7.00
11	16.00	16	1.00	36.00	4.00	.	.
12	17.00	17	4.00	20.00	2.00	5.00	7.00
13	19.00	19	1.00	41.00	4.00	.	.
14	20.00	20	1.00	27.00	3.00	.	.
15	21.00	21	1.00	24.00	2.00	.	.

# Warning !! ... Do data cleansing first

For Example (My Idea)

- identify N/A data form ... 0, 99, blank, N/A
- understandable names and labels
- logic linkage checking

# Let's transform SPSS File in R

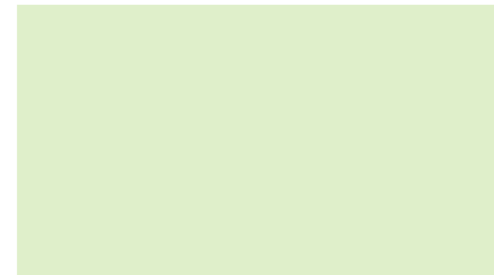
```
> masterdata$Q11_1
<Labelled double>: Q11_1.Time 1st
 [1] 5 5 5 5 6 5 5 5 5 5 5 5 5 6 5 4 6 6 6 6 5 6 5
 [39] 6 5 5 8 6 5 5 6 6 4 2 6 8 5

Labels:
 value          label
 1      6.00 - 8.59  นาฬิกา
 2      9.00 - 11.59 นาฬิกา
 3     12.00 - 14.59  นาฬิกา
 4     15.00 - 16.59  นาฬิกา
 5     17.00 - 19.59  นาฬิกา
 6     20.00 - 22.59  นาฬิกา
 7     23.00 -  5.59  นาฬิกา
 8  ไม่นานอน (ไม่ต้องแสดงการ์ด)
> class(masterdata$Q11_1)
 [1] "haven_labelled"
```

By Using **Haven** package, object's class for SPSS file in R will be "haven\_labelled".

# Let's transform SPSS File in R

So we should  
transform them into  
dataframe  
Objects...



# Why? / How?

- Why? ... it's easier to perform **data wrangling** with dataframe objects.
- How? ... thinking that we will decompose it like 3 sheets of MS Excel file ( or 3 tables of Database ).

# Decomposing SPSS File !!! – 3 Tables

1. Data Structure – Questions
2. Data Structure – Answers
3. Data Input

# Table 1/3 : Data Structure – Questions

	qhead	qdesc
1	Case_ID	Case_ID
2	ID	ID.
3	Check_Group	Check_Group
4	S2	S2.ระบุอายุจริง
5	S2Group	S2Group.S2
6	Q11_1	Q11_1.Time 1st
7	Q12_1	Q12_1.Day 1st
8	Q13_1	Q13. - 1
9	Q10Att1	Q10. - Att1_รู้สึกสะดวกที่มีบริการส่งของถึงบ้าน
10	Q10Att2	Q10. - Att2_ประหยัดเวลากว่าเมื่อเทียบกับการไปเลือกซื้อสินค้าด้..
11	Q10Att3	Q10. - Att3_ไม่ต้องถือของหนักกลับบ้านเอง
12	Q10Att4	Q10. - Att4_ร้านค้าออนไลน์มีสิ่งที่เงินต้องการในขณะที่ร้านค้าใน...
13	Q10Att5	Q10. - Att5_สามารถเปรียบเทียบคุณสมบัติและราคาของสินค้าได้...
14	Q10Att6	Q10. - Att6_สามารถเลือกซื้อสินค้าได้ตลอดเวลา
15	Q10Att7	Q10. - Att7_ราคาสินค้าถูกกว่าซื้อตามร้านค้า

## Fields from SPSS Variable View

- Name
- Label



## Table 2/3 : Data Structure – Answers

	question	choice	desc
53	S2	64	0064
54	S2	65	0065
55	S2Group	2	18-24 ปี
56	S2Group	3	25-35 ปี
57	S2Group	4	มากกว่า 35 ปี
58	Q11_1	1	6.00 - 8.59 นาที
59	Q11_1	2	9.00 - 11.59 นาที
60	Q11_1	3	12.00 - 14.59 นาที
61	Q11_1	4	15.00 - 16.59 นาที
62	Q11_1	5	17.00 - 19.59 นาที
63	Q11_1	6	20.00 - 22.59 นาที
64	Q11_1	7	23.00 - 5.59 นาที
65	Q11_1	8	ไม่แน่นอน (ไม่ต้องแสดงการ์ด)

### Fields from SPSS Variable View

- Name
- Value : Value
- Value : Label

# Table 3/3 : Data Input

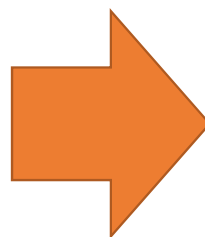
	Case_ID Case_ID	ID ID.	Check_Group Check_Group	S2 S2.สมชายเจริญ	S2Group S2Group.S2	Q11_1 Q11_1.Time 1st	Q12_1 Q12_1.Day 1st	
1	6	6	3	42	4	5	1	
2	7	7	3	27	3	5	1	
3	9	9	3	27	3	5	1	
4	10	10	3	22	2	5	7	
5	11	11	3	18	2	6	8	
6	12	12	3	18	2	5	1	
7	13	13	3	24	2	5	7	
8	26	26	3	25	3	5	7	
9	28	28	3	50	4	5	1	
10	33	33	3	37	4	5	1	
11	34	34	3	42	4	5	1	
12	35	35	3	41	4	5	1	
13	41	41	3	31	3	5	7	
14	42	42	3	27	3	5	6	

Fields from SPSS  
Data View

# Create Fillable Form – Items Selection

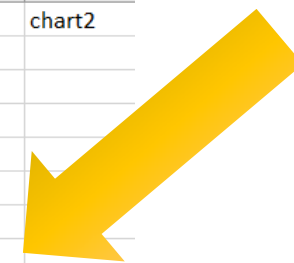
## Blank Form

	A	B	C	D	E	F
1		qhead	qdesc	freqSumm	chart1	chart2
2	1	Case_ID	Case_ID			
3	2	ID	ID.			
4	3	Check_Grc	Check_Grc			
5	4	S2	S2.ระบุการ			
6	5	S2Group	S2Group.S			
7	6	Q11_1	Q11_1.Tim			
8	7	Q12_1	Q12_1.Day			
9	8	Q13_1	Q13. - 1			
10	9	Q10Att1	Q10. - Att1			
11	10	Q10Att2	Q10. - Att2			
12	11	Q10Att3	Q10. - Att3			
13	12	Q10Att4	Q10. - Att4			



## Filled Form

	A	B	C	D	E	F
1		qhead	qdesc	freqSumm	chart1	chart2
2	1	Case_ID	Case_ID			
3	2	ID	ID.			
4	3	Check_Grc	Check_Grc			
5	4	S2	S2.ระบุการ			
6	5	S2Group	S2Group.S			
7	6	Q11_1	Q11_1.Tim	y	y	
8	7	Q12_1	Q12_1.Day	y	y	
9	8	Q13_1	Q13. - 1	y	y	
10	9	Q10Att1	Q10. - Att1	y		y
11	10	Q10Att2	Q10. - Att2	y		y
12	11	Q10Att3	Q10. - Att3	y		y
13	12	Q10Att4	Q10. - Att4	y		y



# Combined Table – From items selected

	question	wvalue	vlabel	PercFreq
1	Q10Att15	1	ค่อนข้างไม่เห็นด้วย	0.17307692
2	Q10Att15	2	เฉยๆ	0.30769231
3	Q10Att15	3	ค่อนข้างเห็นด้วย	0.50000000
4	Q10Att15	4	เห็นด้วยอย่างยิ่ง	0.01923077

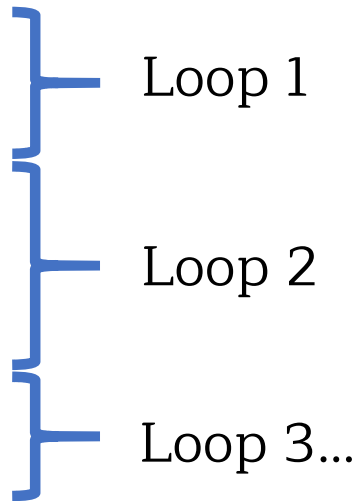
Eg. Making Frequency Table

- Aggregate Frequency data from [Table 3 : data input](#)

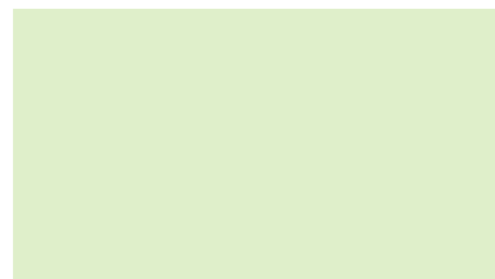
- Look up (join) with Data description from [Table 2 : Data Structure – Answers](#)

# Combined Table – Looping

	question	wvalue	vlabel	PercFreq
19	Q10Att3	1	เฉยๆ	0.07692308
20	Q10Att3	2	ค่อนข้างเห็นด้วย	0.53846154
21	Q10Att3	3	เห็นด้วยอย่างยิ่ง	0.38461538
22	Q10Att4	1	ค่อนข้างไม่เห็นด้วย	0.01923077
23	Q10Att4	2	เฉยๆ	0.17307692
24	Q10Att4	3	ค่อนข้างเห็นด้วย	0.53846154
25	Q10Att4	4	เห็นด้วยอย่างยิ่ง	0.26923077
26	Q10Att5	1	ค่อนข้างไม่เห็นด้วย	0.03846154
27	Q10Att5	2	เฉยๆ	0.21153846

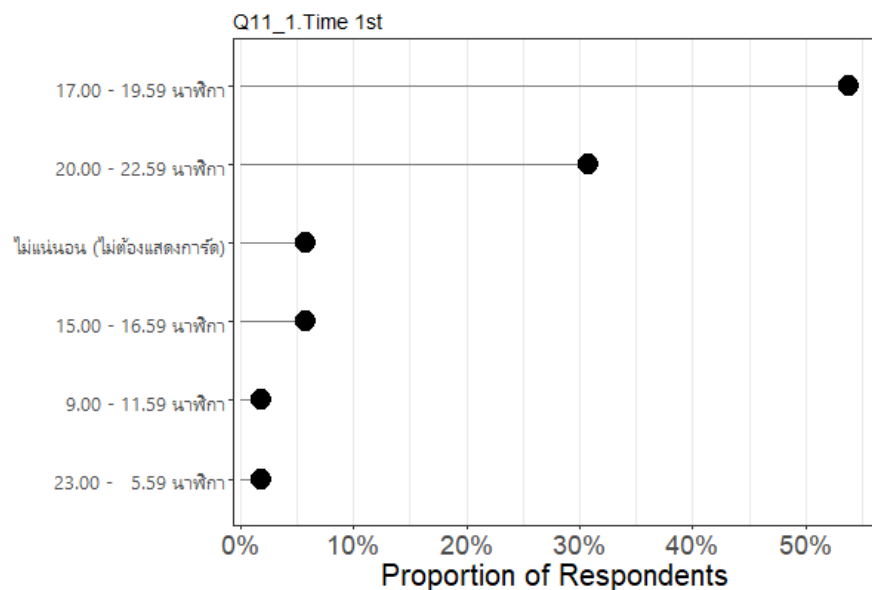


“for loop”  
--- for (){} ---



# DATA AUTOMATION

## Visualization



## Summarization

	question	wvalue	vlabel	PercFreq
1	Q10Att15	1	ค่อนข้างไม่เห็นด้วย	0.17307692
2	Q10Att15	2	เฉยๆ	0.30769231
3	Q10Att15	3	ค่อนข้างเห็นด้วย	0.50000000
4	Q10Att15	4	เห็นด้วยอย่างยิ่ง	0.01923077

