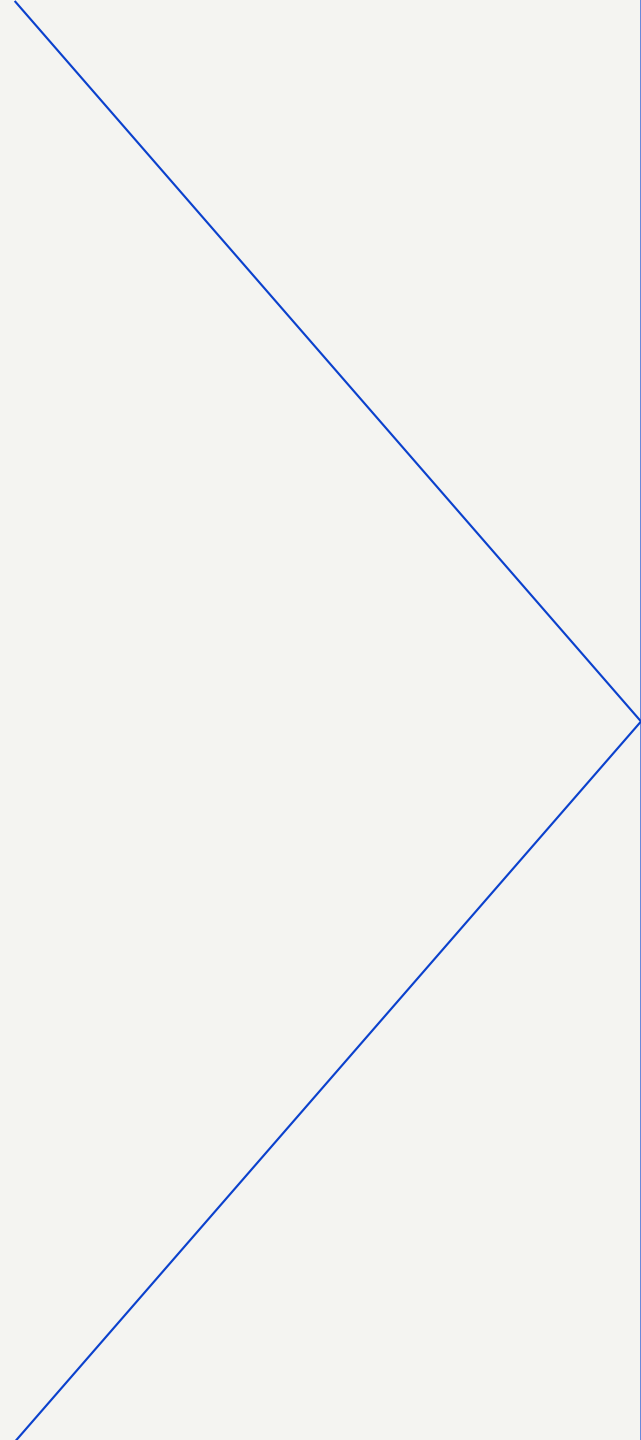


Pixel Plots: Identify Missing Data, See High Level Patterns and Break Axis Constraints

PHUSE EU Connect 2022: DV07

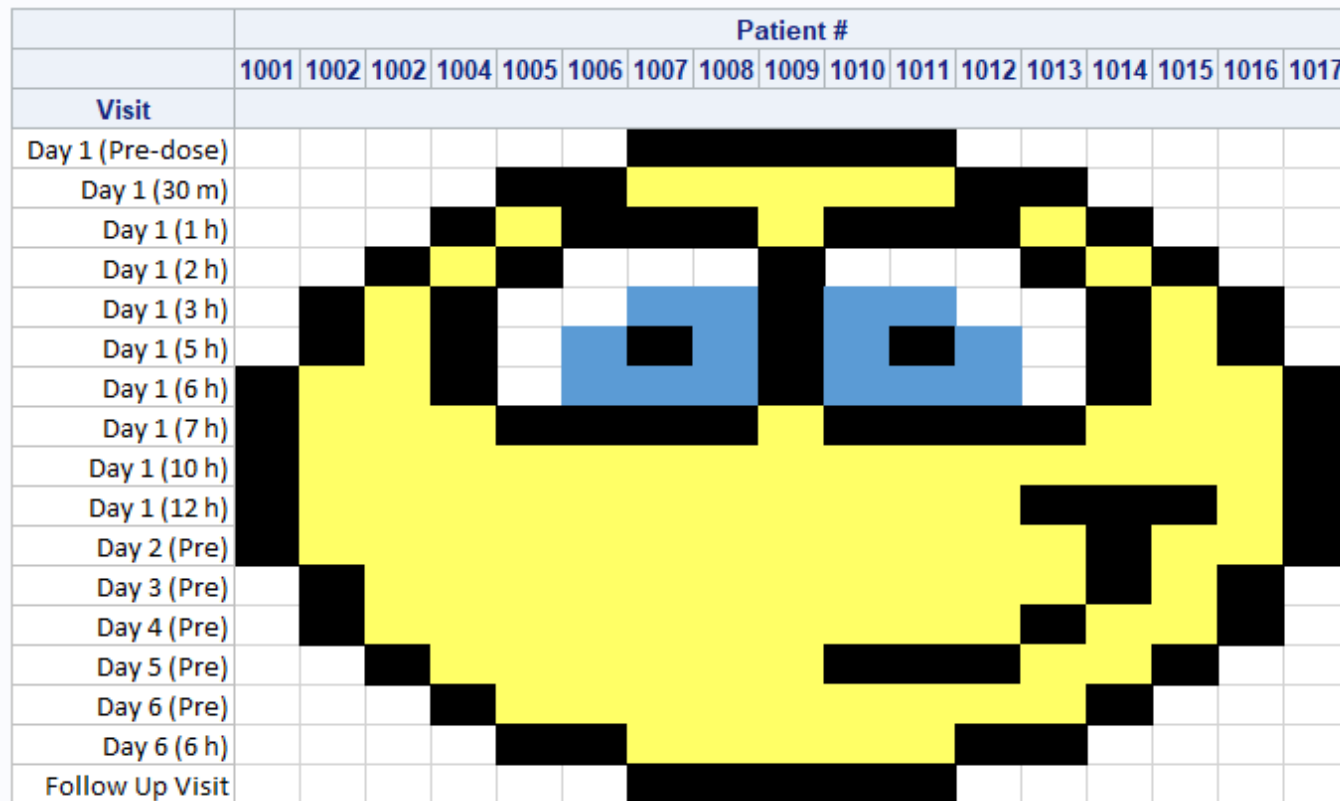
Paul Grimsey, PKPD Data Scientist, Roche Products Ltd, United Kingdom

Table of contents

- 
- A decorative graphic consisting of a vertical blue line and two diagonal blue lines that meet at a point on the vertical line, forming a large right-pointing arrow shape.
1. Introduction
 2. Examples
 3. How to Create the Visualisation
(R and SAS)
 4. Points to Consider
 5. Conclusions

Introduction

What Are Pixel Plots?



- A representation of a 2 dimensional data set
- Each “pixel” represents a different value or values in the data set
- “Pixel” colours are user-defined
- All examples in this presentation are plotting patient and visit / scheduled time point data

Why Are Pixel Plots Useful?

	Patient #																
	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017
Visit																	
Cycle 1 Day 1 (Predose)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 1 Day 1 (Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 1 Day 1 (2 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 1 Day 2 (24 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 1 Day 3 (48 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 1 Day 8	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 1 Day 15	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 2 Day 1 (Predose)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 2 Day 1 (Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 2 Day 1 (2 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 2 Day 2 (24 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 2 Day 3 (48 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 2 Day 8	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 2 Day 15	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 3 Day 1 (Predose)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 3 Day 1 (Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 3 Day 1 (2 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 3 Day 3 (48 H Eoi)	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 3 Day 8	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Cycle 3 Day 15	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
End Of Treatment	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Follow Up	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
# Unshed Visits				1		1											

Key: **green** = PK concentration
black = BLQ
purple = not done

- They allow you to see:-
 - **how much data you have**
 - **where there are gaps in the data**
- **Data issues may also be spotted** e.g. Cycle Day 1 Predose concentration for patient 1002
- By use of colouring **trends can be identified** (see later examples)

This example shows all PK samples received for all patients on a study. Cells are coloured based upon the result (see key below). N.B. this “basic” visualisation uses a single SDTM as the data source

What Else Does it Do?

What Else is Possible?

- **Anything you want!** providing it can be programmed
- Along with identifying what data is received, **trends can be identified by simple colouring** such as:-
 - A patient achieves an efficacy response
 - A patient develops antibodies
 - Unexpected dosing regimens
 - A patient has an AE of interest
- The following slides show some examples that have been used in clinical studies
- Can you think of any other applications?

New Data Visualisation

	Patient #																									
Scheduled Time (hr)	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026
Day 1 (2h)																										
Day 1 (4h)																										
Day 1 (6h)																										
Day 2 (Pre)																										
Day 14 (Pre)																										
Day 28 (Pre)																										
Day 28 (2h)																										
Day 28 (4h)																										
Day 28 (6h)																										
Day 56 (Pre)																										
Day 56 (2h)																										
Day 56 (4h)																										
Day 56 (6h)																										
Day 112 (Pre)																										
Day 196 (Pre)																										
Day 196 (2h)																										
Day 196 (4h)																										
Day 196 (6h)																										
Day 280 (Pre)																										
Day 364 (Pre)																										
Day 364 (2h)																										
Day 364 (4h)																										
Day 364 (6h)																										

Key:

Grey = unchanged

Red = new data

Purple = changed / removed data

- Combines data from the new delivery with the previous delivery to **see what has changed**
- Very **useful for communicating with other functions** that are also analysing the data e.g. dose decision work / ongoing safety monitoring

Basic Visualisation Greying Out Non-Protocol Visits

	Patient #																
Visit	1	2	3	11	12	13	14	15	16	17	18	19	21	22	23	24	25
Cycle 1 Day 1 (Pre)																	
Cycle 1 Day 1 (EOI)																	
Cycle 1 Day 1 (6H)																	
Cycle 1 Day 2 (24H)																	
Cycle 1 Day 3 (48H)																	
Cycle 1 Day 8																	
Cycle 1 Day 8 (Pre)																	
Cycle 1 Day 8 (EOI)																	
Cycle 1 Day 8 (6H)																	
Cycle 1 Day 9 (24H)																	
Cycle 1 Day 15																	
Cycle 2 Day 1 (Pre)																	
Cycle 2 Day 1 (EOI)																	
Cycle 2 Day 1 (6H)																	
Cycle 2 Day 2 (24H)																	
Cycle 2 Day 3 (48H)																	
C2 D8																	
C2 D15																	
Cycle 3 Day 1 (Pre)																	
Cycle 3 Day 1 (EOI)																	
Cycle 3 Day 1 (6H)																	
Cycle 3 Day 2 (24H)																	
Cycle 3 Day 3 (48H)																	
Cycle 3 Day 8																	

Key: **green** = PK concentration
black = BLQ
grey = visits not expected as per protocol

- Removes blank space where visits do not occur per protocol
- Showing blank cells when patient should not take the visit per protocol is misleading
- Need to determine what visits can and cannot be taken from protocol Schedule of Assessments for each patient cohort
- This example shows 3 different patient cohorts in a Phase I dose escalation study

Dosing Visualisation

	Patient #																			
Visit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Day 1	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Week 5	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Week 13	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black
Week 21	Blue	Blue	Blue	Blue	Blue	White	Blue	Blue	Blue	Blue	Black	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Week 29	Black	Black	Black	Black	Black	White	Black	Black	Black	Black	Black	Black	White	Black	Black	Black	White	Black	Black	Black
Week 37	Blue	Blue	Blue	Blue	Blue	White	Blue	Black	White	Blue	White	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Week 45	Black	Black	Black	Black	Black	White	Black	Black	White	Black	Black	White	Black	Black	Black	Black	Black	Black	Black	Black
Week 53	Blue	Blue	Blue	Blue	Blue	White	Blue	Blue	White	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Week 61	Black	Black	Black	Black	Black	White	Black	Black	White	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black
Week 69	White	Blue	Blue	Blue	Blue	White	Blue	Blue	White	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Week 77	Black	Black	Black	Black	Black	White	Black	Black	White	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black
Week 85	Blue	Blue	Blue	Blue	Blue	White	Blue	Blue	White	Blue	White	White	White	White	White	White	White	White	White	White

Key:

Blue = active treatment

Black = placebo treatment

- Subset of an 800 patient phase III study.
- Able to **find missing / unexpected dosing regimens** e.g. patient 11 and use these “unusual” patients to help validate derivations using dosing dates
- **“Unusual” patients can also be communicated to the study team**
- Like the basic visualisation, **this uses a single SDTM**

Anti-Drug Antibody (ADA) Visualisation

				Visit															
				Cycle 1 Day 1	Cycle 1 Day 15	Cycle 2 Day 1	Cycle 2 Day 15	Cycle 3 Day 1	Cycle3 Day 15	Cycle 4 Day 1	Cycle 5 Day 1	Cycle 6 Day 1	Cycle 7 Day 1	Cycle 8 Day 1	Cycle 9 Day 1	Cycle 10 Day 1	End of Treatment	Safety Follow Up	
Cohort	Patient	Country	Dose Group																
A	1001	ESP	0.05MG				ESP												
	1002	ESP	0.05MG																
	1003	BEL	0.05MG																
B	2001	DNK	0.5MG																
	2002	DNK	0.5MG																
	2003	USA	0.5MG																
	2004	USA	0.5MG																
	2005	BEL	0.5MG																
	2006	DNK	0.5MG																
	2007	AUS	0.5MG																
C	3001	CAN	1MG																
	3002	BEL	1MG																
	3003	CAN	1MG																
	3004	ESP	1MG																
	3005	USA	1MG																
D	4001	DNK	2MG																
	4002	BEL	2MG																
	4003	ESP	2MG																
	4004	USA	2MG																
	4005	AUS	2MG																

Key:

Green = ADA negative

Red = ADA positive

- The purpose of this visualisation is to **show when patients become ADA positive**
- A variation of the basic visualisation, showing patients on y-axis and visits on x-axis
- **Combines 2 data sources:** ADA and demography
- This visualisation was requested by Clinical Pharmacologist as was previously hand generating this table for presentation to the project team at each data delivery

Visualisation Showing Missing / Expected Data

- As the basic visualisation but also **shows which visits were taken but no samples are yet received**
- Requires **3 data sources**:-
 - Sample results data e.g. PC
 - Protocol Schedule of Assessments – user defined and created
 - Subject visit dataset SV
- Use the last 2 data sources to determine which expected visits have been taken. Then combine with PC to check if results are present.
- **Can be difficult to maintain** due to protocol changes, or patients taking visits which are not expected as per protocol

	Patient #			
	1	2	3	4
Visit				
Cycle 1 Day 1 (Pre)	Green	Green	Green	Yellow
Cycle 1 Day 1 (2h)	Black	Green	Black	Yellow
Cycle 1 Day 1 (Eoi)	Green	Black	Green	Yellow
Cycle 1 Day 2 (20h)	Green	Green	Green	Yellow
Cycle 1 Day 3 (44h)	Green	Green	Green	Yellow
Cycle 1 Day 8	Green	Green	Green	Yellow
Cycle 2 Day 1 (Pre)			Purple	
Cycle 2 Day 1 (Eoi)			Green	
Cycle 3 Day 1 (Pre)			Black	
Cycle 3 Day 1 (Eoi)			Green	
End Of Treatment	Black	Black		
Fu Until Relapse	Black	Black	Yellow	

Key:

green = PK conc

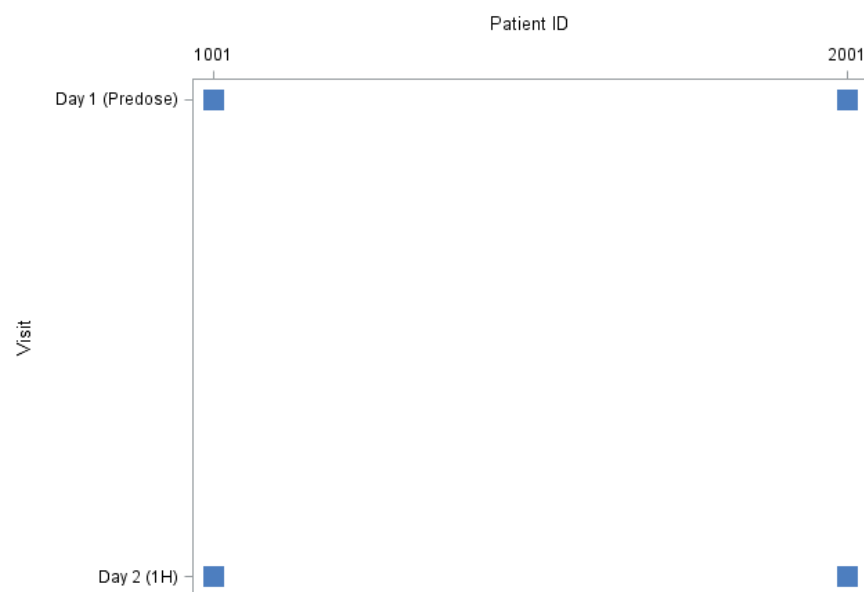
black = BLQ

purple = not done

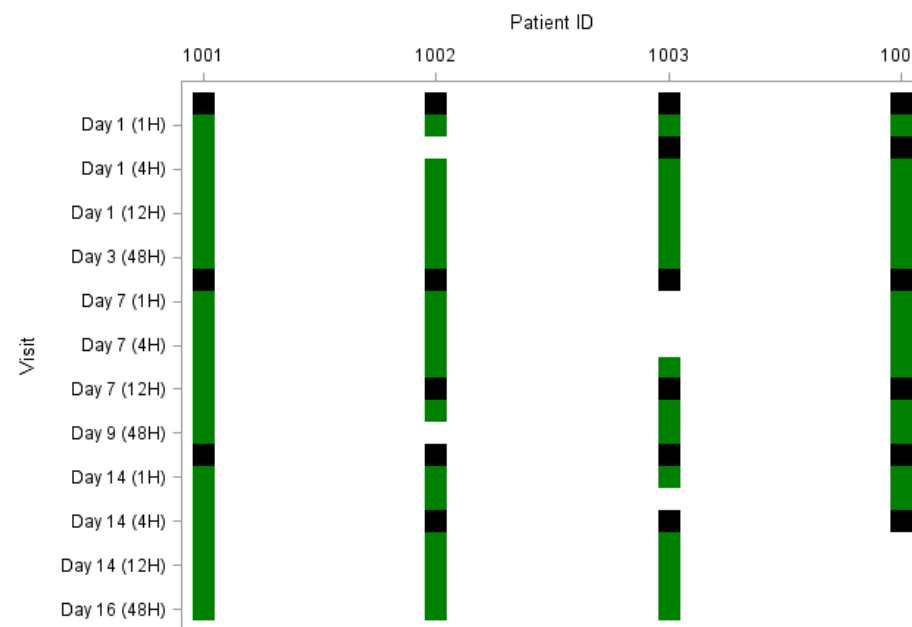
Yellow = visit taken but data not yet received

Overcoming Issues with Static Plots

Static Plots: Issues With Too Little, Or Too Much Data



Example of a small amount of data being plotted



Example of a large amount of data being plotted

- **With few data points the “cells” the gaps between cells are wide.** This can make the visualisation difficult to read
- **With lots of data points the “cells” merge and information can be lost** on the axes (note: in the visualisation on the left the predose, 2H, 8H and 24H time points have been omitted due to the amount of data being displayed)
- Cell size and axes can be manually adjusted but this is not efficient

Export to Excel

[illegible]

34	Day 546 (2h)							
35	Day 546 (4h)							
36	Day 546 (6h)							
37	Day 546 (Day)							

PK

Metabolite 1

ADA

Metabolite 2

+

- **Export to Excel overcomes these issues:-**
 - Cells are always adjacent to one another, do not merge and no truncation of data occurs
 - The zoom function, row and column sliders can be used to navigate the visualisation and **not be limited by axis constraints in static plots**
 - Header row can be frozen programmatically
- Data types can be **split over multiple worksheets** in the same workbook

How To Create The Visualisation

Data Structure for Creating the Visualisation

	1	2	3	4	5
	ID	REPVIS	VISITOR	RESULT	TEST
1	10001	CYCLE 1 DAY 1 (PREDOSE)	2001	B	RO1234567
2	10001	CYCLE 1 DAY 1 (6 HOURS POST DOSE)	2002	R	RO1234567
3	10001	CYCLE 1 DAY 1 (END OF INFUSION)	2005	R	RO1234567
4	10001	CYCLE 1 DAY 2 (24 HOURS POST DOSE)	3003	R	RO1234567
5	10001	CYCLE 1 DAY 3 (48 HOURS POST DOSE)	4004	B	RO1234567
6	10001	CYCLE 1 DAY 8	5000	B	RO1234567
7	10001	CYCLE 1 DAY 15	7000	B	RO1234567
8	10001	CYCLE 2 DAY 1 (PREDOSE)	8001	B	RO1234567
9	10001	CYCLE 2 DAY 1 (END OF INFUSION)	8005	R	RO1234567
10	10001	CYCLE 2 DAY 2 (24 HOURS POST DOSE)	9003	B	RO1234567
11	10001	CYCLE 2 DAY 8	11000	B	RO1234567
12	10001	CYCLE 2 DAY 15	12000	B	RO1234567

The basic dataset creation requires **5 columns**:-

1. **ID** - Numeric patient identifier e.g. from USUBJID
2. **REPVIS** - Combination of scheduled visit and time point (e.g. VISIT and PCTPT)
3. **VISITOR** – numeric ordering of REPVIS (user-defined, or use VISITNUM and PCTPTNUM)
4. **RESULT** – used to define the cell colouring.
5. **TEST** – Name of analyte. Used for splitting over multiple worksheets

This example uses PC but **the data source could be anything you wish to visualise** e.g. LB, EX, etc.

Strategies for Creating the Visualisations in R and SAS

R (and Python)

- **Transpose the dataframe** (see below)
- Define the output to Excel using the **openxlsx** package
- Create a **loop for multiple worksheets**
- Create a **dataframe for the visualisation information** (see next section)

SAS

- Create a **format to display REPVIS using VISORD**
- Data transpose and cell colouring are done in **PROC REPORT**
- Output to Excel using **ODS EXCEL**
- Use **TITLE statements** for information about the visualisation (see next section)

Please see the paper for R and SAS sample code

	REPVIS	VISITOR	TEST	10001	10002	10003	21001	21002	21003	21004
1	CYCLE 1 DAY 1 (PREDOSE)	2001	RO1234567	B	B	B	B	B	B	B
2	CYCLE 1 DAY 1 (6 HOURS POST DOSE)	2002	RO1234567	R	R	R	R	R	R	R
3	CYCLE 1 DAY 1 (END OF INFUSION)	2005	RO1234567	R	R	R	R	R	R	R
4	CYCLE 1 DAY 2 (24 HOURS POST DOSE)	3003	RO1234567	R	R	R	R	R	R	R
5	CYCLE 1 DAY 3 (48 HOURS POST DOSE)	4004	RO1234567	B	R	R	R	R	R	R
6	CYCLE 1 DAY 8	5000	RO1234567	B	R	R	R	B	R	NA

Example of a transposed dataframe in R showing ID as the column headers and RESULT as the column values.

How to Add Other Information to the Visualisation

Every good visualisation should have a title, key and information about where it was created. Here are 2 ways this can be done in the Excel visualisation:

As a separate worksheet

Has the benefit of being able to include lots of information but the downside is that it is not next to the visualisation

	A	B	C	D	E	F	G
1	Key.Values						
2	Study: WP12345						
3							
4	Key:-						
5	Green = Received						
6	Black = BLQ						
7	Purple = Not Done						
8							
9	This visualisation shows each sample per patient as a coloured cell (colour codes given above)						
10	Unscheduled visits show the number of unscheduled samples only.						
11							
12	Program location: \myprojects\WP12345\programs\l_pk_samples.r						

	A	B	C	D	E
1	<study number> - Received PK Samples				
2	Program location: <program location>				
3					
4	Key:				
5	green = PK conc				
6	black = BLQ				
7	purple = changed data				
8	yellow = visit taken but no PK result				
9					
10		Patient #			
11		1	2	3	4
12	Visit				
13	Cycle 1 Day 1 (Pre)				
14	Cycle 1 Day 1 (2h)				
15	Cycle 1 Day 1 (Eoi)				
16	Cycle 1 Day 2 (20h)				
17	Cycle 1 Day 3 (44h)				
18	Cycle 1 Day 8				
19	Cycle 2 Day 1 (Pre)				
20	Cycle 2 Day 1 (Eoi)				
21	Cycle 3 Day 1 (Pre)				
22	Cycle 3 Day 1 (Eoi)				
23	End Of Treatment				
24	Fu Until Relapse				
25					
26					

Above the visualisation

Has the benefit of being next to the visualisation but too many lines of information can detract from it.

Points to Consider

Duplicate Samples

This applies to samples with the same date but different, or identical results). These are handled differently in the R and SAS strategies:

R

- **Must be removed** from R code, otherwise the dataframe transpose will fail
- The sample code will output these for the user to determine which should be kept

SAS

- The visualisation runs but these **samples are plotted as a separate lines**

	Patient #						
	10001	10002	10003	21001	21002	21003	21004
Visit							
C1 D1 (PRE)							
C1 D1 (EOI)							
C1 D1 (6H)							
C1 D2 (24H)							
C1 D3 (48H)							
C1 D8							
C1 D15							
C2 D1 (PRE)							
C2 D1 (EOI)							
C2 D1 (6H)							
C2 D2 (24H)							

Duplicate samples displayed in SAS visualisation.

Unscheduled Visits

Unscheduled visits **do not fit neatly into this type of visualisation as:**

- Patients can have multiple unscheduled visits
- These visits usually have the same name
- They can occur at different dates / time

Here are 2 solutions:

- **remap manually** to include scheduled visit in name e.g. Unscheduled Cycle 1 Day 1
- **total these visits** - this is the solution used in the sample code

	Patient #						
	1026	1027	1028	1076	1077	1078	1079
Visit							
Day 546 (6h)							
Day 644 (Pre)							
Day 728 (Pre)							
Day 728 (2h)							
Day 728 (4h)							
Day 728 (6h)							
130 Ole							
Disc							
# Unsched Visits	4			1	2		

Unscheduled visits are totalled at bottom of the visualisation.

Conclusions and Acknowledgments

Conclusions

- These types of visualisations make it easy to **see what data is present and what data is missing**
- They can be an **effective way of communicating the data with other functions**
- Simple colouring allows **identification of high level trends and can even show data issues**
- The visualisation **can be created simply, using a single data source**, or with more complexity using multiple data sources
- It is **recommended to keep these visualisations simple**, to reduce the amount of maintenance required

	Patient #																
	1001	1002	1002	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017
Visit																	
Day 1 (Pre-dose)																	
Day 1 (30 m)																	
Day 1 (1 h)																	
Day 1 (2 h)																	
Day 1 (3 h)																	
Day 1 (5 h)																	
Day 1 (6 h)																	
Day 1 (7 h)																	
Day 1 (10 h)																	
Day 1 (12 h)																	
Day 2 (Pre)																	
Day 3 (Pre)																	
Day 4 (Pre)																	
Day 5 (Pre)																	
Day 6 (Pre)																	
Day 6 (6 h)																	
Follow Up Visit																	

Acknowledgements

I would like to thank my Clinical Pharmacology colleagues who provided me feedback on how to make these visualisations more user-friendly and provided the initial question that started this work

Acknowledgements

Doing now what patients need next