http://www.researchgate.net/profile/Rebecca_Oyomopito http://au.linkedin.com/pub/rebecca-oyomopito/87/ab0/169

Prioritising indicators from items in big data: An algorithm for an automated, visual approach.

Rebecca Oyomopito PhD

2019

Table of contents

What is program automation? - A cake study

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article

Frog cake

From Wikipedia, the free encyclopedia

The **frog cake** is a dessert in the shape of a frog's head, composed of sponge cake and cream covered with fondant. It was created by the Balfours bakery in 1922, and soon became a popular treat in South Australia. Originally frog cakes were available exclusively in green, but later brown and pink were added to the range. Since then other variations have been developed, including seasonal varieties (such as snowmen and Easter "chicks"). The frog cake has been called "uniquely South Australian",[1] and has been employed in promoting the state. In recognition of its cultural significance, in 2001 the frog cake was listed as a South Australian Heritage Icon by the National Trust of South Australia.

1. Let's say, we want to make a frogcake. We find a recipe that gives good results – this is our standard.

```
Data FrogCake;
Follow the recipe;
Run;
```

2. What if we could vary the number of frogcakes we make in one iteration? Pseudo code for 3 cakes could look like this

```
%Macro MakeCakes (number = );
Data FrogCake;
Follow the recipe;
Do it (number = ) times;
Run;
%Mend;

%MakeCakes (number = 3 );
```

3. What about varying the flavour too?

```
%Macro MakeCakes (number = , flavour = );
Data FrogCake;
Follow the recipe;
flavour = ( );
Do it (number = ) times;
Run;
%Mend;

%MakeCakes (flavour = green, number = 2 );
%MakeCakes (flavour = pink,  number = 2 );
%MakeCakes (flavour = choc,  number = 2 );
```

4. But remember, it all depends on getting the recipe right in the first place!

Recipe by Rebecca Oyomopito

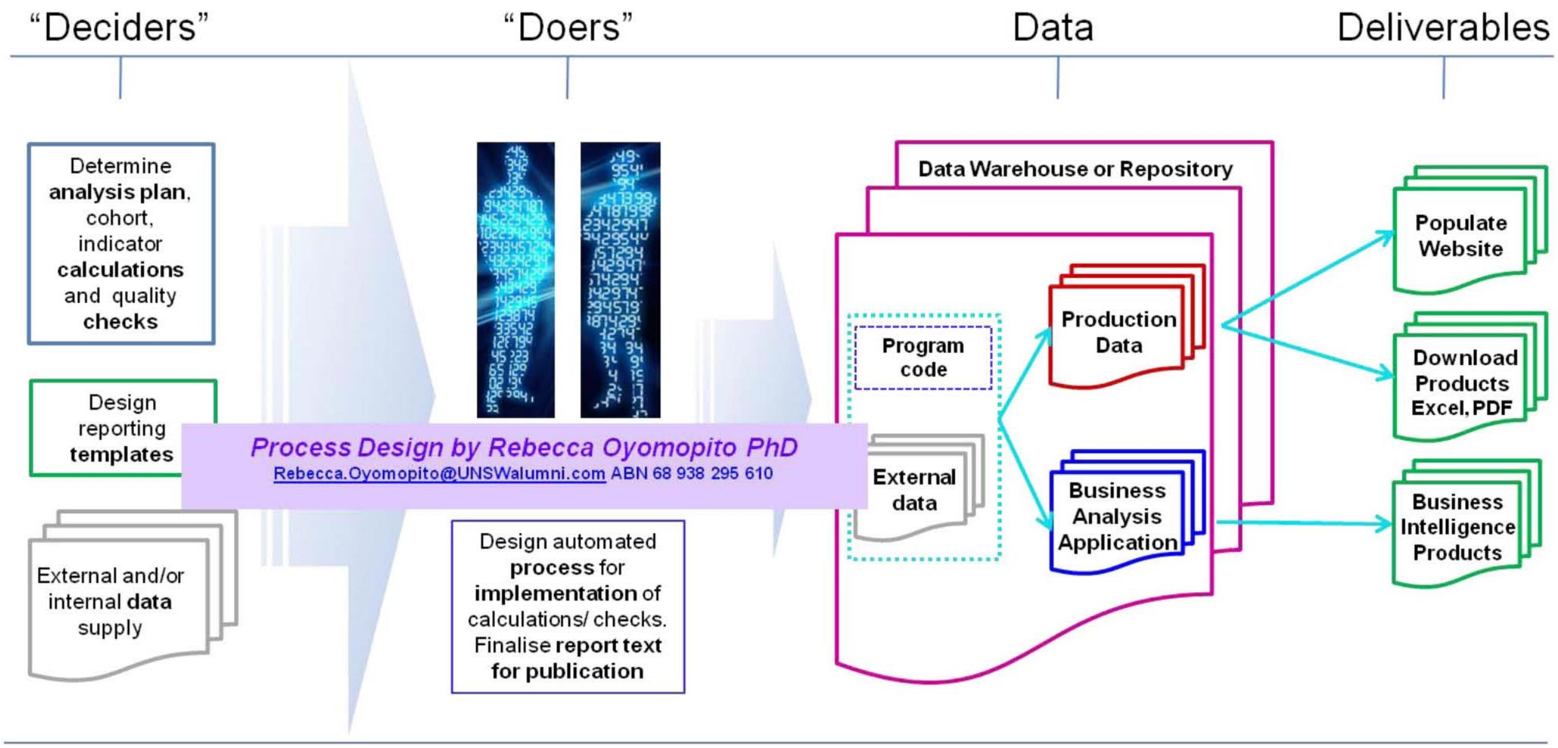
Schematic: An automated reporting pipeline

Dr Rebecca Oyomopito: Australian Business Number (ABN) 68 938 295 610

Roles: Brainstorming facilitation, quality improvement, risk-reduction strategies and technical report.

During 2013-2014, I facilitated brainstorming among 12 colleagues across four teams, designed automated data-to-web processes and collated an operations guide for public reporting detailing risk management and methods to enhance capacity for concurrent reports. I was one of five national pilot testers of a new SAS software platform rollout in an Enterprise Data Warehouse teradata environment for the Australian Department of Health.

CV Reference: National Health Performance Authority (NHPA)



This operational schematic shows activities required to progress publication products through to release.

Graphic 1: Deriving indicators from items

Dr Rebecca Oyomopito: Australian Business Number (ABN) 68 938 295 610

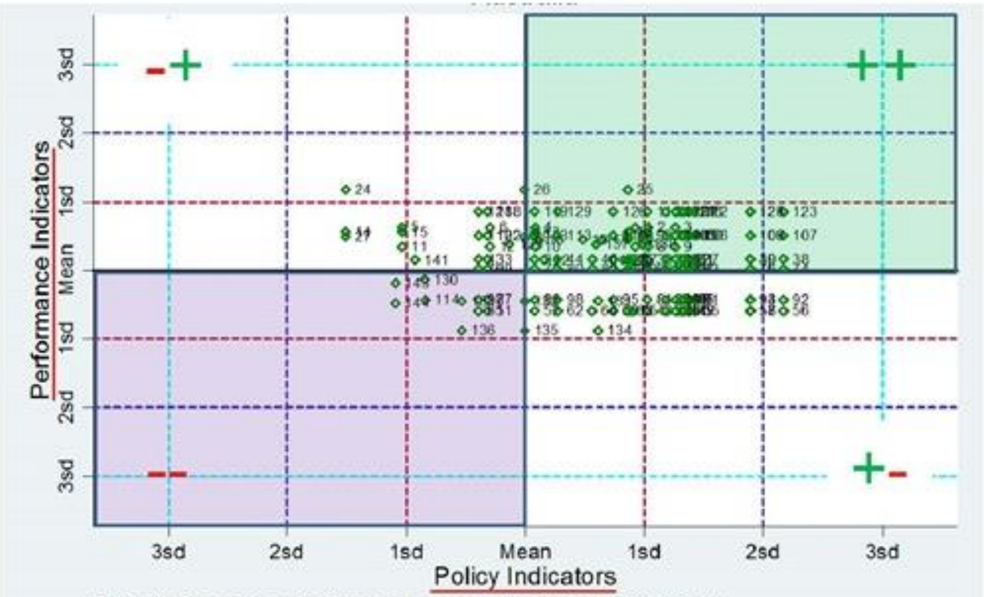
Roles: Concept, multi-country evaluation methodology, novel metrics, graphics and interpretation.

In 2006, to determine which country policy and performance indicators were lower than the OECD average, from concept, I designed methods to assess policy/performance pairs by quadrant and provided definitions and interpretations in an accompanying statistical report.

CV Reference: Organisation for Economic Cooperation and Development (OECD)

OECD countries – Policy and Performance

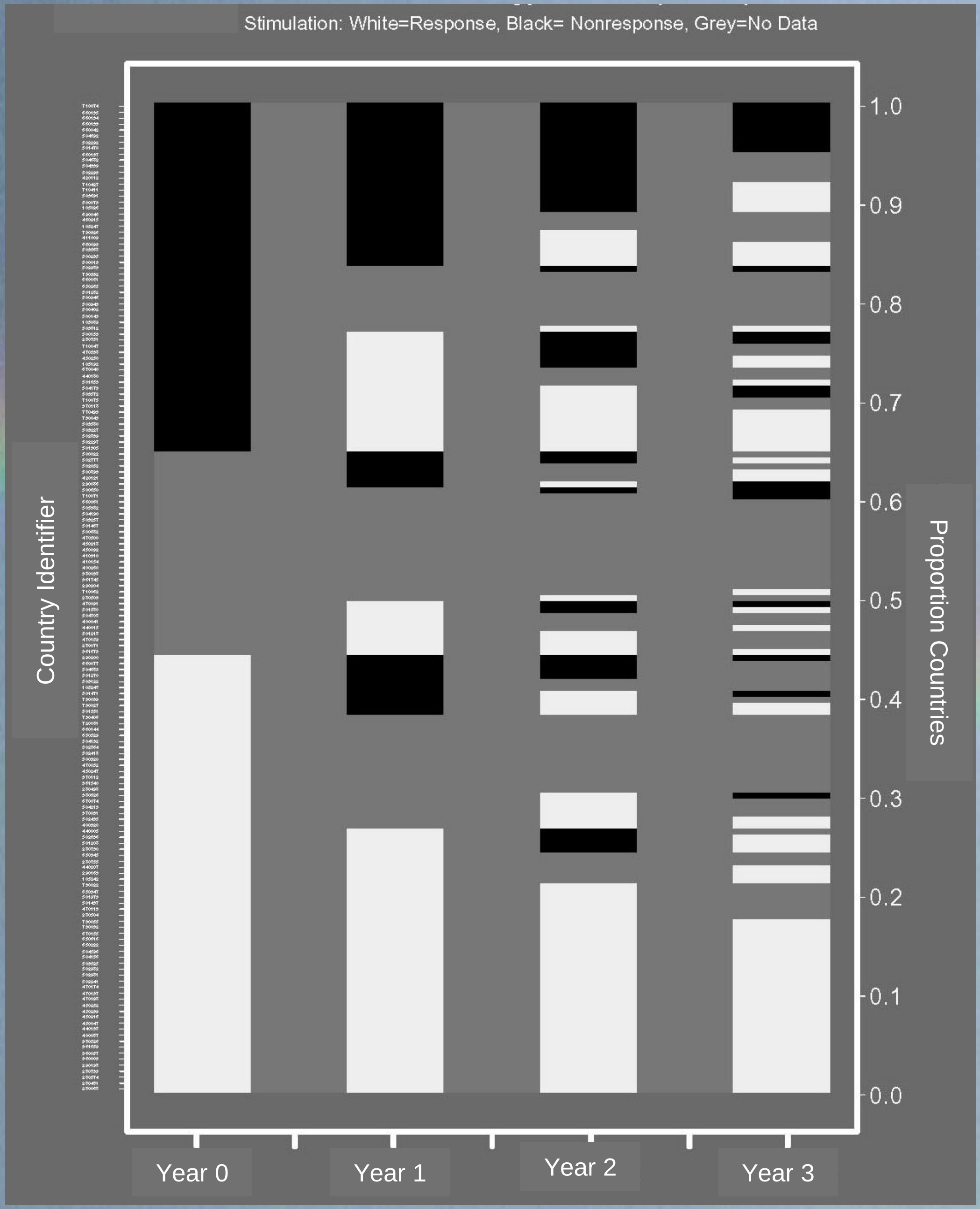
Novel descriptive graphics and evaluation metrics



Note: Performance and policy measures are standardised N(0,1) where: 68%(95%) of area under curve is within 1(2) sds of the mean.

Quadrant presentation of standardised policy and performance indicator pairs facilitates expert interpretation. ++ (--) represents good (poor) scores on both indicators. Poor scoring and discordant pairs (-+, +-) may be investigated by referencing the accompanying indicator definitions and metrics in tabular report format (not shown). For non-normally distributed data, a normalised median approach is preferred.

Graphic 2: Indicator response over time



Year 0	COUNT	PERCENT
Response	73	44.24
Missing Data	34	20.61
NonResponse	58	35.15

Year 1	COUNT	PERCENT
Response	73	44.24
Missing Data	49	29.7
NonResponse	43	26.06

Year 2	COUNT	PERCENT
Response	69	41.82
Missing Data	58	35.15
NonResponse	38	23.03

Year 3	COUNT	PERCENT
Response	66	40
Missing Data	79	47.88
NonResponse	20	12.12

The graphic shows that some countries below the required indicator threshold (in black) at the start of data collection baseline (Year 0), achieved response (white) by Year 3.

However, a more than double increase in missing data (grey), as shown in tables, from 21% to 48%, overshadowed the positive message.

Post algorithm example analysis for innovation indicators

Ten of the innovation indicators load onto Factor 1 with a cut-off value for the correlation between the indicator and this factor of 0.7 (Table 5.A1.3, identifies these variables with a * in the Factor 1 column). Considering the nature of the variables, they appear to reflect *"knowledge development"*. Four other innovation indicators load onto Factor 2 (see Table 5.A1.3, variables identified with a * in the Factor 2 column). These indicators mostly appear to reflect *"knowledge application"*.

On average the Netherlands ranks 9th out of the OECD 20 countries for the indicators of "knowledge development" (Table 5.A1.4); lowest ranks for individual indicators range from 16-20 depending on available data. The Netherlands does particularly well on EPO high-tech patent applications (Item 2.3.1) and public R&D expenditure as a share of GDP (Item 2.1) but scores below average on the proportion of the population with tertiary education (Item 1.2) and business expenditures on R&D as a percentage of GDP (Item 2.2);

Table A1.3. Factor loadings

Item	Loading	Factor 1	Loading	Factor 2
1.1 S&E graduates (‰ of 20 – 29 years age class)	0.53		0.06	
1.2 Population with tertiary education (% of 25 – 64 years age class)	0.78	*	-0.29	
1.3 Participation in life-long learning (% of 25 – 64 years age class)	0.73	*	-0.56	
1.4 Employment in medium-high and high-tech manufacturing (% of total workforce)	0.19		0.64	
1.5 Employment in high-tech services (% of total workforce)	0.88	*	-0.17	
2.1 Public R&D expenditures (% of GDP)	0.89	*	0.19	
2.2 Business expenditures on R&D (% of GDP)	0.90	*	0.29	
2.3.1 EPO high-tech patent applications (per million population)	0.85	*	0.19	
2.3.2 USPTO high-tech patents granted (per million population)	0.87	*	0.35	
3.1 SMEs innovating in-house (% of all SMEs)	-0.02		0.80	*
3.2 SMEs involved in innovation co-operation (% of all SMEs)	0.95	*	-0.04	
3.3 Innovation expenditures (% of total turnover)	-0.07		0.82	*
3.4 SMEs using non-technological change (% of all SMEs)	-0.33		0.74	*
4.1 Share of high-tech venture capital investment	0.35		0.25	
4.2 Share of early stage venture capital in GDP	0.89	*	-0.13	
4.3.1 Sales of 'new to market' products (% of total turnover)	0.12		0.57	
4.3.2 Sales of 'new to the firm but not new to the market' products (% of total turnover)	-0.07		0.86	*
4.4 Internet access	0.68		-0.21	
4.5 ICT expenditures (% of GDP)	0.63		0.08	
4.6 Share of manufacturing value-added in high-tech sectors	0.89	*	0.22	

Source: European Commission (2004b), European Innovation Scoreboard 2004 Database; own calculations.

Table A1.4. Rankings of OECD 20 countries for innovation items that load on "Knowledge Development"

Country	Rank Item 1.2	Rank Item 1.3	Rank Item 1.5	Rank Item 2.1	Rank Item 2.2	Rank Item 2.3.1	Rank Item 2.3.2	Rank Item 3.2	Rank Item 4.2	Rank Item 4.6	Factor 1 Item Average Rank
Finland	3	7	3	2	2	1	3	1	2	3	2.7
Sweden	8	1	1	3	1	3	4	3	1	8	3.3
USA	1	n.a.	n.a.	4	6	6	1	n.a.	4	4	3.4
Japan	2	n.a.	n.a.	7	3	9	2	n.a.	n.a.	7	5.0
Iceland	10	3	2	1	6	8	5	5	7	n.a.	5.2
Switzerland	9	2	7	11	4	4	6	6	5	1	5.5
Denmark	4	6	4	8	7	7	7	2	3	9	5.7
UK	6	5	5	13	12	10	10	12	6	5	8.4
Netherlands	12	8	11	6	14	2	9	8	11	11	9.2
France	15	12	6	5	11	11	11	9	9	6	9.5
Belgium	7	10	8	15	9	12	12	7	10	10	10.0
Norway	5	4	10	10	15	15	15	4	8	14	10.0
Germany	14	14	13	9	8	6	8	10	14	12	10.8
Ireland	11	9	9	19	16	13	13	n.a.	13	2	11.7
Austria	17	11	12	12	13	14	14	11	15	13	13.2
Luxembourg	18	13	14	20	10	16	18	n.a.	n.a.	19	16.0
Spain	13	15	16	17	17	18	17	16	16	16	16.1
Italy	20	16	15	16	18	17	16	15	18	15	16.6
Portugal	19	18	18	14	19	20	20	13	12	17	17.0
Greece	16	17	17	18	20	19	19	14	17	18	17.5

Source: European Commission (2004b), European Innovation Scoreboard 2004 Database; own calculations.

Publications arising:

OECD (2006), OECD Economic Surveys: *Netherlands*, Volume 2006/2, OECD, Paris, prepared for the Committee by D Carey, E Ernst, J Theisens and **R Oyomopito**.
http://www.oecd.org/document/56/0,3746,en_2649_34111_45424120_1_1_1_1,00.htm

Carey, D, Ernst, E, **Oyomopito, R** and Theisens, J (2006), "Strengthening innovation in the Netherlands: making better use of knowledge creation in innovation activities", *OECD Economics Department Working Papers*, No 479.
http://www.oecd.org/LongAbstract/0,3425,en_2649_34117_36133255_1_1_1_1,00.html

Reference Extract for R Oyomopito PhD:

Economics Department of the OECD and she was the statistician responsible for the Netherlands, among other countries. She provided statistical analysis that formed the backbone of the in-depth chapter in the Netherlands Economic Survey and subsequent working paper that we prepared on innovation policy. This analysis identified two factors that were associated with innovation for a large number of OECD countries, enabling us to highlight where strengths and weaknesses lay for the Netherlands and hence what were the priorities for reform. Dr Oyomopito showed great initiative and technical competence on this project. She was pleasant to work with, enthusiastic and reliable. Interest in the report was such that extracts were presented at a seminar at the Central Planning Bureau in the Netherlands in December 2005. As you know,

Outcome map: Indicator utilisation

Conclusions

- The algorithmic code automates visual tools and tabular reports to help prioritise meaningful indicators from many items.
- Methods are applicable to economic, health, innovation and digital transformation data.
- Useful for large numbers of entities as found in multi-country, multi-region or geospatial studies.
- Analysis prioritisation, monitoring change and evaluating utilisation are facilitated using minimal resources in a timely fashion.

Future directions

- Promote algorithm in areas where meaningful indicators need to be derived from many items for a large number of entities.
- Leverage the algorithm for big data applications *e.g.* in deriving national indicators for regional health system entities and geospatial boundaries.
- Collaborate with Stakeholders and Key Partners to develop supplementary modules.