

Dissemination - Vorbereitungsmaterial

Maria Molnar

11. Juni 2016

im Rahmen des Masterseminars
Spezielle Themen der Sozial- und Wirtschaftsstatistik

Sommersemester 2016

Institut für Statistik
Ludwig-Maximilians-Universität München

Inhaltsverzeichnis

1	Einführung	3
2	Definition	3
2.1	Unterschiedliche Formen	3
3	Standards	3
3.1	IMF - Special Data Dissemination Standard	3
4	UNECE - Making Data Meaningful	4
4.1	Making Data Meaningful Part 1 - A guide to writing stories about numbers . . .	4
4.2	Making Data Meaningful Part 2 - A guide to presenting statistics	5
4.2.1	Wahrnehmung	5
4.2.2	Tabellen	6
4.2.3	Diagramme	7
4.2.4	Karten	7
4.2.5	Neue Techniken	8
4.3	Making Data Meaningful Part 3 - A guide to communicating with the media . .	9
5	Prospect Theory	10
	Literatur	12

1 Einführung

Das Ziel dieser Seminararbeit ist eine kurze Einführung in die wichtigsten Aspekte der Dissemination zu geben und eine relativ neue Theorie über die menschliche Wahrnehmung von Wahrscheinlichkeiten vorzustellen.

2 Definition

Dissemination - Veröffentlichung und Verbreitung statistischer Daten und Ergebnisse [1]

2.1 Unterschiedliche Formen

Es stehen unterschiedliche Medien zur Verfügung, wenn es um die Verbreitung von statistischer Daten geht. Dank der letzten Jahrzehnten hat sich die Palette durch zahlreiche elektronische Möglichkeiten erweitert.

- Digital: Webseiten, CD-ROM, PDF, E-Book
- Papierform: Jahrbücher, Berichte, Pressemitteilungen, Zeitungsartikeln
- TV, Radio

3 Standards

Ziele von Standards:

- Einheitliche Entwicklung, Produktion und Verbreitung von Statistiken,
- Bereitstellung von zuverlässigen und vergleichbaren Statistiken,
- hohe Qualität zu sichern.

Zahlreiche bekannte Standards, Beispiel:

3.1 IMF - Special Data Dissemination Standard

Internationaler Währungsfonds - International Monetary Fund: Sonderorganisation der Vereinten Nationen.

Ihre Aufgaben [2]:

- Überwachung: die Wechselkurspolitik der Mitgliedsländer wird im Rahmen einer umfassenden Analyse der allgemeinen wirtschaftlichen Lage und der wirtschaftspolitischen Strategie eines jeden Mitgliedslandes beurteilt,

- Finanzhilfen: Kredite und Darlehen, die Mitgliedsländern in Zahlungsbilanzschwierigkeiten zur Unterstützung ihrer wirtschaftspolitischen Anpassungs- und Reformmaßnahmen gewährt wird,
- Technische Hilfe: Fachkenntnissen und Unterstützung, die den Mitgliedsländern in mehreren breitgefassten Bereichen zur Verfügung gestellt wird, wie zum Beispiel Gestaltung und Umsetzung der Geld- und Fiskalpolitik; Institutionsaufbau und Erfassen und Aufbereitung statistischer Daten.

Ziele der Special Data Dissemination Standard [3]:

- Datentransparenz der Mitgliedstaaten zu verbessern,
- Entwicklung von statistischen Systemen zu fördern,
- angemessene Informationen über die wirtschaftliche Lage der Mitgliedstaaten bereitzustellen.

4 UNECE - Making Data Meaningful

Die Wirtschaftskommission für Europa der Vereinten Nationen hat mehrere praktische Anleitungen veröffentlicht, in denen man Leitlinien dazu findet, wie man am besten Texte, Tabellen, Diagramme und zusätzliche Informationen für die Öffentlichkeit zusammenstellt. Es geht nicht um Regeln, sondern nur um Vorschlägen und Beispiele.

In diesem Abschnitt werden diese Anleitungen und ihre wichtigsten Aspekte vorgestellt, diese Anleitungen wurden unter dem Titel Making Data Meaningful veröffentlicht.

4.1 Making Data Meaningful Part 1 - A guide to writing stories about numbers

- Statistiken selbst sind für die Leser nur Zahlen. Wenn man ihre Aufmerksamkeit wecken will, muss man diese in einer leicht verständlichen und unterhaltsamen Weise vorstellen.
- Leser können die Daten und Ergebnisse viel leichter merken und verstehen, wenn es eine Geschichte darum gebaut wird.
- Artikel so schreiben, wie ein Journalist
 - Bericht, wie eine Erzählung
 - wichtigste Punkte: Themen, erster Absatz, Schlagzeile, einfache Sprache
- Zielgruppe und Kommunikation mit den Zielgruppen - wegen dem Internet anders, als früher
- Bewertung der Auswirkung von den Artikeln: Verfolgen, Analyse der Zeitung, Webseite, Besucheranzahl, was könnte verbessert werden, neue Themen, Weiterbildung

4.2 Making Data Meaningful Part 2 - A guide to presenting statistics

Neben dem richtig formulierten Text hilft die geeignete Visualisierung am meisten beim Verstehen der Daten. Deswegen sollte es bei der Dissemination der Daten eine zentrale Rolle spielen, dass man die beste Lösung dafür findet. Mit einer gut gewählten Darstellung kann man auch solche Zusammenhänge entdecken, die sonst verborgen bleiben würden.

Die meisten Leser haben schon Erfahrungen mit den üblichen Visualisierungen, wie Balkendiagramme und Karten, sie verstehen diese viel leichter, als nur Zahlen. Natürlich kann es vorkommen, dass die Daten doch falsch interpretiert werden. Das kann man vorbeugen, indem man nicht zu komplizierte Grafiken erstellt und diese ausschließlich die nötigsten Informationen enthalten. Außerdem soll man auch auf die kulturellen Unterschiede achten, diese können ebenfalls Fehlinterpretationen zur Folge haben, zum Beispiel wenn man solche Farben verwendet, die in den verschiedenen Ländern andere Bedeutung haben. [5]

In diesem Abschnitt geht es darum, was alles man bei der Visualisierung beachten soll, welche Techniken zur Verfügung stehen und wie man mit diesen umgeht, wenn man Daten effektiv darstellen möchte.

4.2.1 Wahrnehmung

Bevor man zu den einzelnen Darstellungstypen kommt, ist es wichtig noch kurz einen Aspekt der menschlichen Wahrnehmung zu erwähnen.

Auf Abbildung 1 sieht man zwei orangefarbene Kreise, die von weiteren Kreisen umgeben sind. Der rechte Kreis scheint größer zu sein, als der linke Kreis. Sie sind aber gleich groß, wenn man die blauen Kreise abdeckt, sieht man es sofort. Es ist nur eine optische Täuschung, die Ebbinghaus-Täuschung genannt wird. [5] Dieses Beispiel zeigt, dass es im Gehirn die verschiedenen Objekte, die man sieht, in Beziehung gesetzt werden. Bei der Darstellung statistischer Daten sollte man auch dieses Phänomen beachten, damit mögliche Fehlinterpretationen vermieden werden.

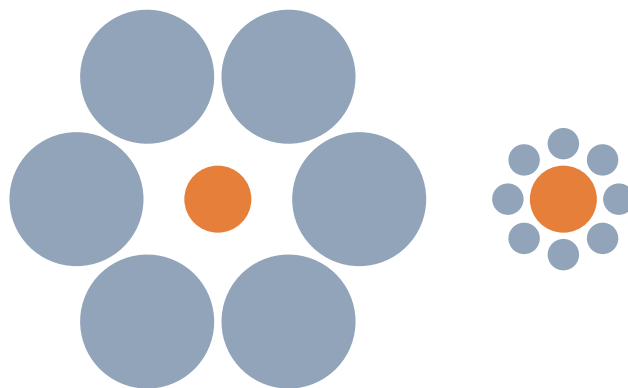


Abbildung 1: Ebbinghaus-Täuschung [8]

4.2.2 Tabellen

Wenn man einen guten Überblick über die Daten verschaffen und zu viel konkrete Daten im Text vermeiden möchte, bietet es sich Tabellen zu benutzen. In der Tabelle sollte nur eine Teilmenge der Daten in kurzer, knapper Weise dargestellt sein, damit die Leser sie schnell durchschauen und verstehen können.

Die Tabellen sollten so gestaltet sein, dass sie auch ohne die ursprüngliche Publikation verständlich sind und eventuell auch in anderen Publikationen eingefügt werden können. Dazu dienen die Metadaten, die jede Tabelle enthalten soll. Solche Informationen sind zum Beispiel die Quelle der Daten oder der Titel der Tabelle. Neben dem Inhalt erleichtert auch selbst die Struktur der Tabelle das Verstehen. [5]

Es gibt fünf Elemente, die in einer guten Tabelle unerlässlich sind:

- die **Überschrift**, in dem eindeutig beschrieben ist, was die Tabelle enthält, er soll die Fragen was, wo und wann beantworten;
- die **Spaltenköpfe** und **Zeilenköpfe**, in denen die Daten in den einzelnen Spalten bzw. Zeilen und die dazugehörigen Metadaten beschrieben sind;
- die **Fußnoten** mit den noch nötigen Informationen;
- und die **Quelle** der Daten.

Tabelle 1 [5] stellt ein gutes Beispiel dar. Einerseits enthält sie die oben genannten fünf Elemente, andererseits ist auch die Formatierung ideal. Die Tabelle ist nicht breiter, als nötig, die Spalten sind gleich breit, die Zahlen sind nach dem Dezimalpunkt ausgerichtet und es gibt keine leeren Zellen, obwohl es auch fehlende Werte gibt. Stattdessen wurde n/a eingetragen und in der Fußnote erklärt.

Tabelle 1: **Share of total energy consumption, by sector (in percent)**; Ireland, 1980-2003

	1980	1985	1990	2000	1995	2002	2003
Transport	27.8	27.9	28.2	31.1	36.8	39.5	39.1
Residential	31.1	33.9	30.4	27.6	24.3	23.7	24.0
Industry	31.5	27.2	23.9	22.1	21.4	19.5	18.8
Agriculture	n/a ¹	n/a ¹	3.5	3.7	3.1	2.9	2.8
Services	9.6	11.0	14.0	15.5	14.4	14.4	15.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

¹ Data on energy consumption for the agricultural sector was not collected until 1990.

Source: Department of Public Enterprise, Ireland

Wenn es in einer Publikation mehrere Tabellen gibt, sollte man noch außerdem darauf achten, dass sie immer einheitlich gestaltet sind und man die eventuellen Abkürzungen konsequent benutzt.

4.2.3 Diagramme

Besonders starke Werkzeuge für Visualisierung sind die Diagramme. Mit den richtig gewählten Diagrammen kann man die Daten schnell und leicht darstellen und vergleichen bzw. Zusammenhänge und Trends erkennen.

Zuerst soll man aber überlegen, ob es wirklich Sinn macht, die Daten mit Diagrammen darzustellen. In manchen Fällen, wenn es zum Beispiel zu viele oder zu wenige Werte vorhanden sind, bieten Diagramme nicht unbedingt die beste Lösung.

Wenn man in einer Publikation schon für das Verwenden von Diagrammen entschieden hat, sollte man Einiges immer im Hinterkopf haben. Erstens soll man die Zielgruppe und ihre Kenntnisse über die Fragestellung kennen. Es soll außerdem klar sein, was alles man mit den Daten zeigen kann und will. Wie viele mögliche Aussagen kann man treffen? Welche möchte man von diesen Aussagen zeigen, ist es zum Beispiel ein Vergleich von den Daten oder steht ein zeitlicher Trend im Mittelpunkt? Damit man die beste Lösung findet, sollte man zuerst die unterschiedlichen Möglichkeiten ausprobieren und dann die am meisten geeignete Darstellung wählen.

In diesem Abschnitt werden die unterschiedlichen Diagramme kurz vorgestellt.

- Säulen- und Balkendiagramm

Wenn es um Häufigkeitsverteilungen von diskreten Variablen mit relativ wenigen Ausprägungen geht, greift man am besten zu Säulen- und Balkendiagrammen. Dabei werden die Ausprägungen auf der horizontalen Achse abgetragen und die Häufigkeiten auf der vertikalen Achse oder umgekehrt. Eine andere Variante ist das gestapelte Säulendiagramm, wobei relative Häufigkeiten dargestellt werden. Die ganze Säule bildet den Gesamtwert ab und die Häufigkeiten sind als Flächen dargestellt.

- Liniendiagramm

Liniendiagramme eignen sich am besten dazu, Entwicklung und Trends im Laufe der Zeit darzustellen.

- Kreisdiagramm

Eine unter den Statistikern nicht so anerkannte Veranschaulichung ist das Kreisdiagramm. Mit diesem Diagramm kann man am besten eine diskrete Variable und ihre Verteilung darstellen, aber es ist nur begrenzt anwendbar. Wenn es zu viele Ausprägungen gibt, dann ist das Kreisdiagramm schon unübersichtlich. Außerdem kann es schwierig sein, die Kreissegmente zweier Kreisdiagramme oder in dem gleichen Kreisdiagramm zu vergleichen. Diese Schwierigkeit kann man aber mit Beschriftung der Kreissegmente überwinden.

- Streudiagramm

Wenn es um die Darstellung gemeinsamer Messwerte von stetigen Variablen geht, greift man am besten zu den Streudiagrammen. Die Messwerte werden in diesem Fall in ein zweidimensionales Koordinatensystem eingetragen, wodurch eine Punktwolke entsteht. Mit Hilfe dieser Darstellung kann man einen Eindruck darüber gewinnen, ob und wie stark die Variablen zusammenhängen.

4.2.4 Karten

Außer den Messwerten selbst weisen oft statistische Daten auch einen räumlichen Bezug auf. Wenn man diese Eigenschaft ausnutzt, kann man solche Zusammenhänge und Muster erkennen,

die ohne eine räumliche Darstellung versteckt bleiben würden. Solche Daten sind zum Beispiel die Zensusdaten, bestimmte Daten aus der Wirtschaft, Feinstaubmessungen, also alle möglichen Arten von Daten, die man vorstellen kann.

So einen räumlichen Bezug darzustellen eignen sich am besten die Karten. Wenn sie richtig benutzt werden, kann man mit ihnen auf einen Blick eine große Menge von Informationen vermitteln und das Verstehen dieser Informationen erleichtern.

Dank der großen technischen Entwicklung in den letzten Jahrzehnten stehen solche Werkzeuge zur Verfügung, mit denen die raumbezogenen Daten gespeichert, verwaltet, verarbeitet, reorganisiert, modelliert, analysiert und dann in Kartenform visualisiert werden können. Unter anderem kennt man die Geographischen Informationssysteme (GIS) und die Kartographischen Informationssysteme (CIS). Die Daten werden dabei digital erfasst und durch EDV gestützte Verfahren bearbeitet. Es stehen zahlreiche kostenpflichtige und kostenlose Systeme zur Verfügung, wie zum Beispiel ArcGIS von dem Softwarehersteller ESRI oder sogar das Programm R. R kennt man in erster Linie als statistische Software, aber es kann auch als GIS benutzt werden, die verschiedenen geographischen Pakete ermöglichen das.

Man unterscheidet zwei Arten von Karten, die topographischen Karten und die thematischen Karten. Die topographischen Karten dienen der Orientierung im Gelände und der Ausmessung von geometrischen Details, und die thematischen Karten stellen ein bestimmtes Merkmal dar, wie zum Beispiel die Bevölkerungsdichte. Außerdem spricht man einerseits von statischen Karten, die von den Anwendern nicht verändert werden können, andererseits von interaktiven Karten, die animiert oder individuell gestaltbar sind.

Wie bei den Diagrammen gibt es unterschiedliche Arten von Karten bzw. thematischen Karten. Wenn man in einer Publikation diese benutzen will, muss man sich natürlich zuerst überlegen, welche sich am besten zur Datenveranschaulichung eignen. Die folgenden drei thematischen Karten [5] unterscheidet man:

- die **Choroplethenkarte** oder **Flächenkartogramm**, wobei die Gebiete im Verhältnis zur Verteilungsdichte des dargestellten Merkmals eingefärbt werden;
- die **Karte mit Punkten** (dot map), in der die Lage und Dichte einer Population oder eines Phänomens mit Hilfe von Punkten dargestellt werden;
- und die **Karte mit proportionalen Symbolen**, wobei absolute Werte mit Hilfe von Symbolen, zum Beispiel mit Kreisen veranschaulicht sind. Die Größe der Symbole ist dabei proportional zum absoluten Wert des dargestellten Merkmals.

Bei der Erstellung und Verwendung von Karten gelten ähnliche Leitlinien wie bei den Tabellen und Diagrammen. Sie sollten so einfach sein, wie es geht, damit die Anwender alles schnell und leicht verstehen können. Außerdem sollten die Karten auch ohne die Publikation selbst verständlich sein.

4.2.5 Neue Techniken

- Neue Techniken machen die Daten und deren Analyse für die Anwender viel interessanter
- Internet - Kommunikation in beide Richtungen, Anwender können selbst Tabellen und Diagramme erstellen, analysieren und über ihre Entdeckungen diskutieren

- mögliche Fehlinterpretationen
- deswegen brauchen die statistischen Organisationen klare Richtlinien, wie sie die neuen Techniken verwenden

Beispiele der neuen Techniken:

- Dynamische Darstellungen
- Animationen und Video (Gapminder)
- Web 2.0 und Gemeinschaften
- Sparklines, tag clouds

4.3 Making Data Meaningful Part 3 - A guide to communicating with the media

Kommunikation mit den Medien, Zusammenarbeit mit den Journalisten

- Grundsätze
- Verstehen den Bedarf der Medien, damit sie effektiv in der Datenveröffentlichung teilnehmen können
- Kooperation statistischer Organisationen und Medien: spezielle Personel dafür einstellen
- Neue Form der Medien: Social Media - dynamische Werkzeuge, twitter, facebook, blog, etc.
- Zielgruppe hat sich verändert - Selbsbedienung der Anwender - Metadata noch wichtiger
- Rolle der Medien auch anders - Vermittler, die die rohen Daten umwandeln
- Webseite - wie sie richtig gestaltet werden sollte - auch für die Medien
 - einfache Sprache
 - inverted pyramid style
 - Feedback von den Anwendern beachten
 - Veränderungen - Anwender aufmerksam machen
 - Online Press Room - speziell für die Medien ausgearbeiteter Bereich auf der Webseite
 - Kommunikation mit den Journalisten
- Pressemitteilungen schreiben - wer, wie, wann mitteilen
- Pressekonferenz
- Monitoring, Überwachung der Medien
- Medientraining für Statistiker bzw. Journalisten
- Medienkrise behandeln

5 Prospect Theory

Wie man es auch in den vorigen Kapiteln gesehen hat, ist es wichtig zu verstehen, wie die Menschen denken, wenn man ihnen über statistische Daten berichten möchte. Allein schon die Formulierung kann ihre Meinung und eventuell ihre Entscheidungen beeinflussen. Die Wissenschaftler Daniel Kahneman und Amos Tversky haben unter anderem dieses Phänomen erkannt und in ihrer Theorie zum ersten Mal in 1979 beschreibt.

- Entscheidungstheorie - Verhalten von Individuen, Gruppen, Organisationen in Entscheidungssituationen und mit den Konsequenzen dieser Entscheidungen, interdisziplinär
- In den 40er Jahren - Wirtschaftswissenschaftler: der Mensch entscheidet immer rational, das Ziel ist, das erwartete Nutzen zu maximieren (Homo oeconomicus)
- In den 70er Jahren - Empirische Studien: Hunderte von Experimenten mit Glücksspielen - die Teilnehmer sollten Entscheidungen treffen. Es wurde nachgewiesen, dass das Verhalten der Menschen in Risikosituationen nicht rein rational beschrieben werden kann.
- Wissenschaftler Daniel Kahneman und Amos Tversky stellten basierend auf den erwähnten Experimenten ein neues Modell auf, die sogenannte **Prospect Theory** bzw. neue Erwartungstheorie. Dieses Modell bietet eine Alternative zum rein rationalen Modell, in der die Rationalität durch Kognitive Verzerrungen beeinflusst wird.
- Prospect Theory
 - Wie entscheiden Individuen unter Risiko? (Entscheiden unter Risiko bedeutet, die Wahrscheinlichkeiten für das Eintreten der möglichen Umweltzustände zu kennen.)
 - Wie bewerten Individuen erwartete Gewinne und Verluste?
Assymetrisch, Verluste werden stärker gewichtet, als Gewinne
 - Entscheidungsprozess besteht aus zwei Stufen, **editing** (Beschreibung) und **evaluation** (Bewertung)
 - * **editing**: Analyse, Reformulieren und Ordnen der möglichen Resultate
Verfahren in der editing-Phase: coding, combination, segregation, cancellation, simplification, detection of dominance
 - * **evaluation**: Nutzen zu den möglichen Resultaten zuordnen und Alternative mit dem höchsten Nutzen wählen
Einfachste Formel in der evaluation-Phase mit zwei möglichen Resultaten:

$$V(x, p; y, q) = \pi(p) \cdot v(x) + \pi(q) \cdot v(y), \quad (1)$$

wobei V der zugeordnete Wert; x, y die möglichen Resultate und p, q ihre jeweiligen Eintrittswahrscheinlichkeiten sind. $\pi(p)$ ist eine Funktion, die die Wahrscheinlichkeit p gewichtet (Abbildung 2), und $v(x)$ ist eine sogenannte Wertfunktion, die dem jeweiligen Resultat einen Wert bzw. Nutzen in Relation zu einem Referenzpunkt zuordnet (Abbildung 3), der vorher in der editing-Phase bestimmt wurde.

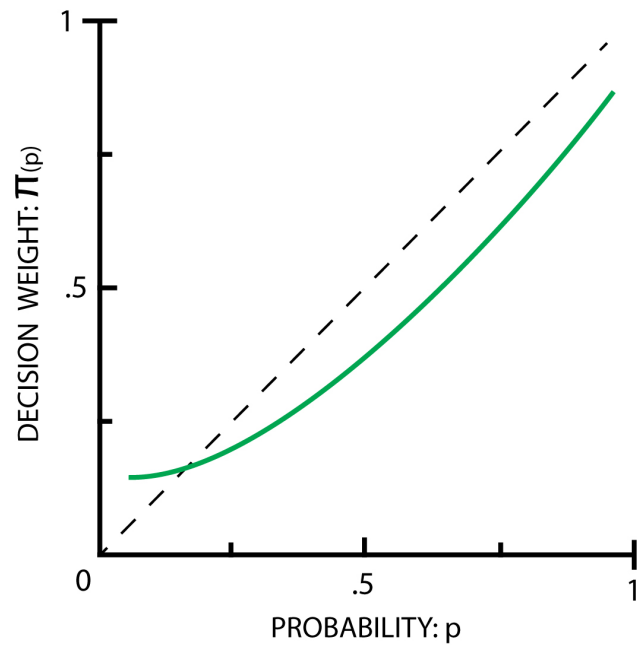


Abbildung 2: Gewichtungsfunktion $\pi(p)$ [9]

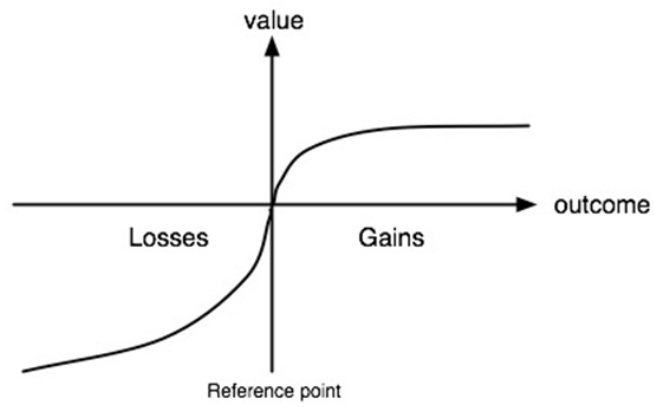


Abbildung 3: Wertfunktion $v(x)$ [10]

Literatur

- [1] <https://stats.oecd.org/glossary/detail.asp?ID=3004>
- [2] <https://www.imf.org/external/np/exr/facts/deu/glanced.htm>
- [3] <http://dsbb.imf.org/images/pdfs/sddsguide.pdf>
- [4] http://www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part1_English.pdf
- [5] http://www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part2_English.pdf
- [6] http://www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part3_English_Print.pdf
- [7] http://www.unece.org/fileadmin/DAM/stats/publications/2013/Making_Data_Meaningful_4.pdf
- [8] https://en.wikipedia.org/wiki/Ebbinghaus_illusion
- [9] <http://www.policonomics.com/wp-content/uploads/Prospect-theory-2.jpg>
- [10] http://www.princeton.edu/~kahneman/docs/Publications/prospect_theory.pdf