

Höchst erstaunliche Graphen

HERMAN CALLAERT, DIEPENBEEK, BELGIEN

übersetzt aus Teaching Statistics, volume 22, number 1, S. 25-27, von Nadine Warmuth, Konstanz

Zusammenfassung: *Erzählt wird eine warnende Geschichte über schlechte Angewohnheiten in graphischen Darstellungen.*

Einführung

Lernen, zumindest auf dem beschreibenden Niveau, wie man die Information einer Graphik versteht, sollte einfach sein. Tatsächlich haben wir alle eine Menge Erfahrung dadurch, daß wir im täglichen Leben einer Menge „aufsehenerregender“ Darstellungen ausgesetzt sind. Dieser Artikel handelt von einem Graphen, der einige schlechte Eigenschaften hat. Er könnte als Beispiel dafür dienen, die Studenten darauf aufmerksam zu machen, kritisch auf jede Datendarstellung zu schauen, der sie in zukünftiger Arbeit begegnen.

Ob einfach oder kompliziert, Graphen werden überall gezeigt und einige sind weit entfernt von einer optimalen Darstellung. Es ist höchst erstaunlich, wie viele neue graphische Darstellungen den Regeln entsprechen, die in H. Wainers Aufsatz

(1994) formuliert sind (Anmerkung: Es wäre klug, erst den Titel des Aufsatzes zu lesen, bevor man stolz ankündigt, daß zumindest die eigenen Graphen Wainers Regeln entsprechen). Studenten zu lehren, kritisch auf Graphen zu schauen, wird ihnen nicht nur bei der Interpretation, sondern auch bei der Konstruktion von graphischen Darstellungen helfen.

Ein faszinierender Graph

Arbeitskosten können von verschiedenen Perspektiven her analysiert werden. Eine von vielen interessierenden Größen mißt den „Brutto-Verdienst“. Daten für diese Variable sind erhältlich für Arbeiter in der Industrie und sie werden zusammengefaßt in einer Zahl pro Land der EU. Die Studie bezieht sich auf 1993, was auch erklärt, warum der Verdienst in EUR ausgedrückt wird. Es gibt 16 Datenpunkte, da für den Zweck der Studie „alte“ und „neue“ Länder in Deutschland als separate Objekte behandelt werden. Der zugehörige Graph ist in Abbildung 1 gezeigt.

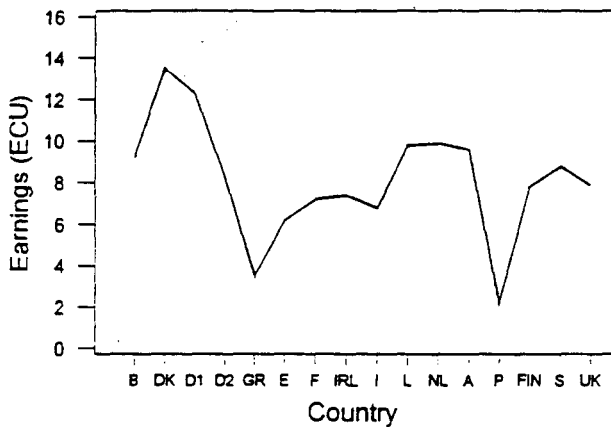


Abb. 1: Bruttostundenlohn von Arbeitern der Industrie 1993 für 15 Staaten der EU.

Die Abkürzungen für Mitgliederstaaten der EU sind:

B Belgien, DK Dänemark, D1 Deutschland „Alte Länder“, D2 Deutschland „Neue Länder“, GR Griechenland, E Spanien, F Frankreich, IRL Irland, I Italien, L Luxemburg, NL Niederlande, A Österreich, P Portugal, FIN Finnland, S Schweden, UK Großbritannien

Ein kurzes Experiment

Unter Verwendung eines Overheadprojektors zeigte ich diesen Graphen einer Gruppe von Studenten. Ich erwähnte den vollen Namen aller Abkürzungen auf der Abzisse. Ich richtete ihre Aufmerksamkeit auch darauf, daß die Länder in alphabetischer Reihenfolge ihres Namens in der Landessprache erscheinen. Das erklärt auch, warum Finnland (welches Suomi auf Finnisch heißt) seinen Platz zwischen Portugal (Republica Portuguesa) und Schweden (Konungariket Sverige) hat. Dann bat ich sie, sich auf Finnland zu konzentrieren und auf Länder, die den Brutto-Verdienst betreffend ähnlich sind. Kurz darauf nahm ich den Graphen weg und bat nun die Studenten, die vier „nächsten Nachbarn“ von Finnland aufzuschreiben. Ich drängte sie außerdem dazu, andere Aspekte zu erwähnen, die ihnen am Graph aufgefallen sind (so z. B. Gruppen von Ländern mit außergewöhnlich niedrigen oder außergewöhnlich hohen Löhnen). Das Ergebnis dieser Untersuchung war ziemlich enttäuschend und der Graph schien nicht sehr hilfreich für das Herausfinden von Informationen über Löhne in EU zu sein.

Als ich später das Experiment anhand des Graphen aus Abbildung 3 mit einer anderen Gruppe von Studenten wiederholte waren Antworten der Studenten erheblich genauer.

Auf die Spitze getrieben

Abbildung 1 ist ein Beispiel für einen (nicht ungewöhnlichen) Typ von Graphen, der geeignet ist, größere Verbesserungen vorzunehmen. Die Datenreihe besteht aus 16 Punkten, jeder trägt die Informationen über den Brutto-Verdienst (kontinuierliche Variable) und das Land (nominale Variable). Die verbundenen Punkte nehmen eine Kontinuität in der Variablen, die auf Abzisse abgetragen ist, an. Das ist bei den Ländern sicherlich nicht der Fall. Abbildung 1 zeigt z. B. eine deutliche Zunahme auf der Verbindungslinie zwischen Portugal und Finnland. Aber es gibt keinen Grund dafür, warum die Länder, benannt und sortiert nach ihren englischen Namen, die Repräsentation mehr (oder weniger) informativ machen. Englische Namen würden Finnland zwischen Dänemark und Frankreich einordnen, auf einen fallenden Teil des Graphen, im Gegensatz zum bisher wachsenden Verlauf des Graphen zwischen Portugal und Schweden. Wenn eigentlich keine natürliche Ordnung der Ländernamen möglich ist (es sind nur Zeichen, die ein Land repräsentieren), dann ist jede Sortierung genauso vernünftig wie jede andere auch. Warum also konstruiert man nicht einen Graphen, den wir alle gut kennen und dessen Form leicht einzuprägen ist?

Die ursprüngliche Abbildung 1 ist nur eine Möglichkeit, die Information in 16 Datenpunkten zu illustrieren. Ohne die Art der Darstellung zu verändern, kann man zwischen 20.922.789.888.000 äquivalenten Alternativen wählen. Warum nicht zu einer gedämpften Oszillation, wie in Abbildung 2 gezeigt, übergehen? Abgesehen von den scharfen Ecken, liefert $8 + 6e^{-0,17x} \sin(2,58x + 1,6)$ eine gute Anpassung.

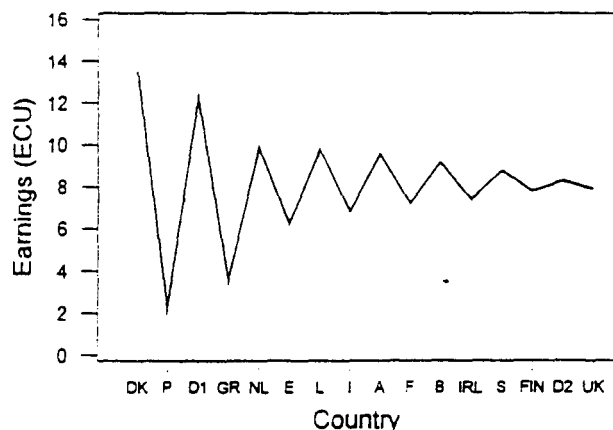


Abb. 2: Dieselbe Information dargestellt mit demselben Typ von Graph, nach einer teilweisen Umordnung der Länder

Nach meinen Erfahrungen versuchen Studenten, die einmal die Äquivalenz der beiden Abbildungen 1 und 2 und den Unsinn der beiden entdeckt haben, weitere Situationen mit „gedämpften Oszillationen“ in Zeitungen, Zeitschriften und Büchern zu finden. Auch wenn sie selbst graphische Darstellungen konstruieren, wollen sie nicht in die Falle „gedämpfter Oszillationen“ tappen.

Eine weitere Abbildung

Wenn man glaubt, daß die Information in den 16 Datenpunkte einer graphischen Repräsentation bedarf, so kann man etwas aus Clewards sehr gutem Buch (1994) lernen. Er erklärt, wie verbundene Liniensegmente die Information über relative lokale Änderungsraten visualisieren können. Seine Bilder auf den Seiten 246/247 zeigen, daß wir für unsere Daten eine Punkt-Darstellung auswählen sollten. Sie zeigen auch, daß das alphabetische Prinzip „Österreich zuerst (Austria first)“ (eine der Regeln von Wainer zur schlechten Darstellung von Daten) unsere Visualisierung der Verdienste erheblich verschlechtert.

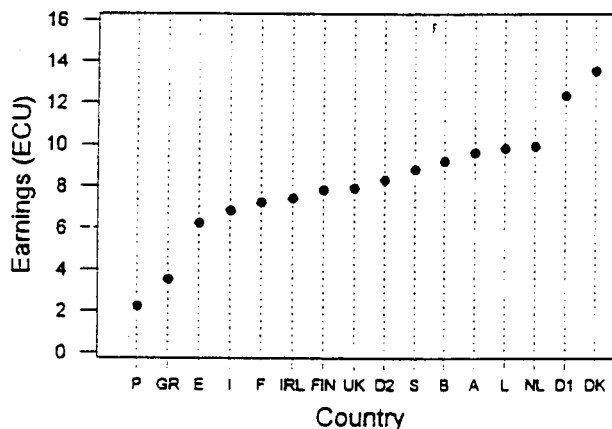


Abb. 3: Dieselbe Information in der Reihenfolge steigender Verdienste

In Abbildung 3 ist die quantitative Information vom kleinsten zum größten geordnet. Gruppen von Ländern mit niedrigen (oder hohen) Verdiensten können leicht erkannt werden. Finnlands „dichteste Nachbarn“ zu finden, ist nun leicht.

Quelle

Der Graph in Abbildung 1 ist auf Seite 197 der 4. Auflage von „Europa in Zahlen“ (1995), herausgegeben von Eurostat, zu finden. Für das Buch wird im Internet mit „Eine Veröffentlichung, die alle wesentlichen sozio-ökonomischen Informationen, die für ein gutes Verständnis der EU gebraucht werden, enthält“ geworben. Es ist weit verbreitet, nicht teuer (15 Euro) und in 9 verschiedenen Sprachen erhältlich. Eurostat ist das statistische Amt der EU.

Literatur

- Cleveland, W.S. (1994): The Elements of Graphing Data (überarbeitete Ausgabe). Summit, NJ: Hobart Press
- Europa in Zahlen (4. Auflage) (1995): Luxemburg: Amt für Veröffentlichungen der EU
- Wainer, H. (1984): How to display data badly. In: The American Statistician, 38, S. 137-47.

Adresse des Autors

Herman Callaert
 Limburgs Universitair Centrum
 Diepenbeek
 Belgium

e-mail: herman.callaert@luc.ac.be