

Tek Değişkenli ve Çok Değişkenli Tablolar ve Grafikler

Umut Al

umutal@hacettepe.edu.tr

Plan

- ❑ İlgili kavramlar
- ❑ Tablo ne zaman kullanılır?
- ❑ Grafik nasıl üretilir?
- ❑ Örnekler
- ❑ Dikkat edilmesi gerekenler

Değişken

- ❑ Variable
- ❑ Değişebilen, yani en az iki değer alabilen her şey
- ❑ Belirli özelliklerin mantıksal seti
 - ❑ Özellik: Erkek - kadın
 - ❑ Değişken: Cinsiyet
- ❑ Bağımlı - bağımsız
- ❑ Sürekli - süreksiz
- ❑ Nicel - nitel

Bağımlı - Bağımsız Değişken

□ Bağımlı

- Açıklanan değişken; sonuç; araştırmacıyı rahatsız eden, açıklanması istenen durum

□ Bağımsız

- Açıklayan değişken; neden

□ Örnekler:

- Sigara * kanser
- Sıcak hava * dondurma tüketimi
- Faiz oranı * hisse senedi fiyatı

Sürekli - Süreksiz Değişken

- ❑ Sürekli
 - ❑ İki değer arasında başka bir değeri bulunabilen değişkenler
- ❑ Süreksiz
 - ❑ Tam sayılarla ifade edilirler
- ❑ Örnekler:
 - ❑ Sürekli => Ağırlık
 - ❑ Süreksiz => Üniversite türleri (devlet üniversitesi, vakıf üniversitesi)

Nicel - Nitel Değişken

- ❑ Nicel

- ❑ Sayılarla ifade edilebilen değişkenler

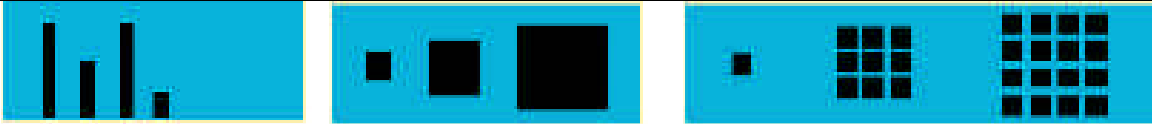
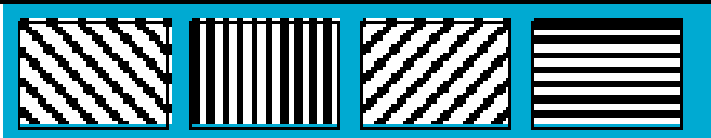
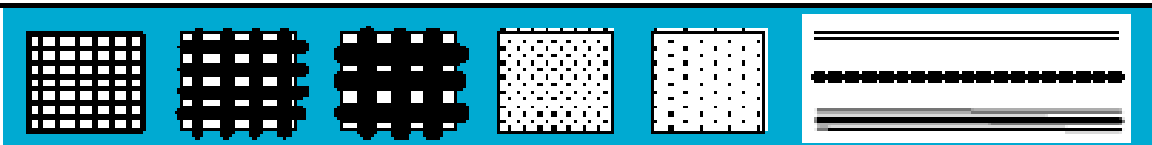
- ❑ Nitel

- ❑ Sembollerle ifade edilebilen değişkenler

- ❑ Örnekler:

- ❑ Nicel => Ağırlık, boy uzunluğu
 - ❑ Nitel => Medeni durum, cinsiyet

Görsel Değişkenler

Bertin's Original Visual Variables	
Position changes in the x, y location	
Size change in length, area or repetition	
Shape infinite number of shapes	
Value changes from light to dark	
Colour changes in hue at a given value	
Orientation changes in alignment	
Texture variation in 'grain'	

Ölçek Türleri

- ❑ Sınıflama
- ❑ Sıralama
- ❑ Eşit aralıklı
- ❑ Oranlı

Ölçek Türleri

❑ Sınıflama

- ❑ Erkek - kadın
- ❑ Başarılı - başarısız

❑ Sıralama

- ❑ Çoktan aza doğru sıralama
- ❑ Çok memnun - memnun - memnun değil - hiç memnun değil
- ❑ Yarışlar => birimler arasındaki aralıklar eşit değil
- ❑ Sıfır değerinin bir anlamı yok

Ölçek Türleri

❑ Eşit aralıklı

- ❑ Kullanıcıların kütüphaneden aldıkları ödünç kitap sayısı (1, 3, 5...)
- ❑ Sıfır gerçek sıfır değil, yokluk ifade etmiyor => suyun donma ve kaynama noktaları => sıcaklık ölçmede termometre kullanımı => iki nokta arası 100 eşit parçaya bölünmüş

❑ Oranlı

- ❑ Kullanıcıların kütüphanede kalış süreleri (iki saat, dört saat...)
- ❑ Gerçek sıfır noktası (ses şiddeti, ağırlık ölçümü)

Tek Değişkenli Tablo

- ❑ Tanımlayıcı istatistiklerin sunumu
- ❑ Olgular tek bir değişkene göre ne şekilde dağılmış

Tablo 1. Bölüm öğrencilerinin yaş gruplarına göre dağılımı

Yaş Grubu	N	%
17-21	88	31
22-26	149	53
27-31	33	12
32 ve üzeri	12	4
Toplam	282	100

Çapraz Tablo

- ❑ Bir değişkenin başka bir değişken üzerindeki etkisi
- ❑ Bağımsız değişken satırda, bağımlı değişken sütunda yer almalı

Tablo 2. Kullanıcıların cinsiyetlerine göre kütüphane kullanımı

Cinsiyet	Kullanıyor	Kullanmıyor	Toplam
Erkek	158	29	187
Kadın	115	29	144
Toplam	273	58	331

Satır ve Sütun Yüzdelерinin Kullanımı

- ❑ Satır ve sütun yüzdesi verirken neden bahsediyoruz?

Tablo 2. Kullanıcıların cinsiyetlerine göre kütüphane kullanımı

Cinsiyet	Kullanıyor	Kullanmıyor	Toplam
Erkek	158	29	187
Kadın	115	29	144
Toplam	273	58	331

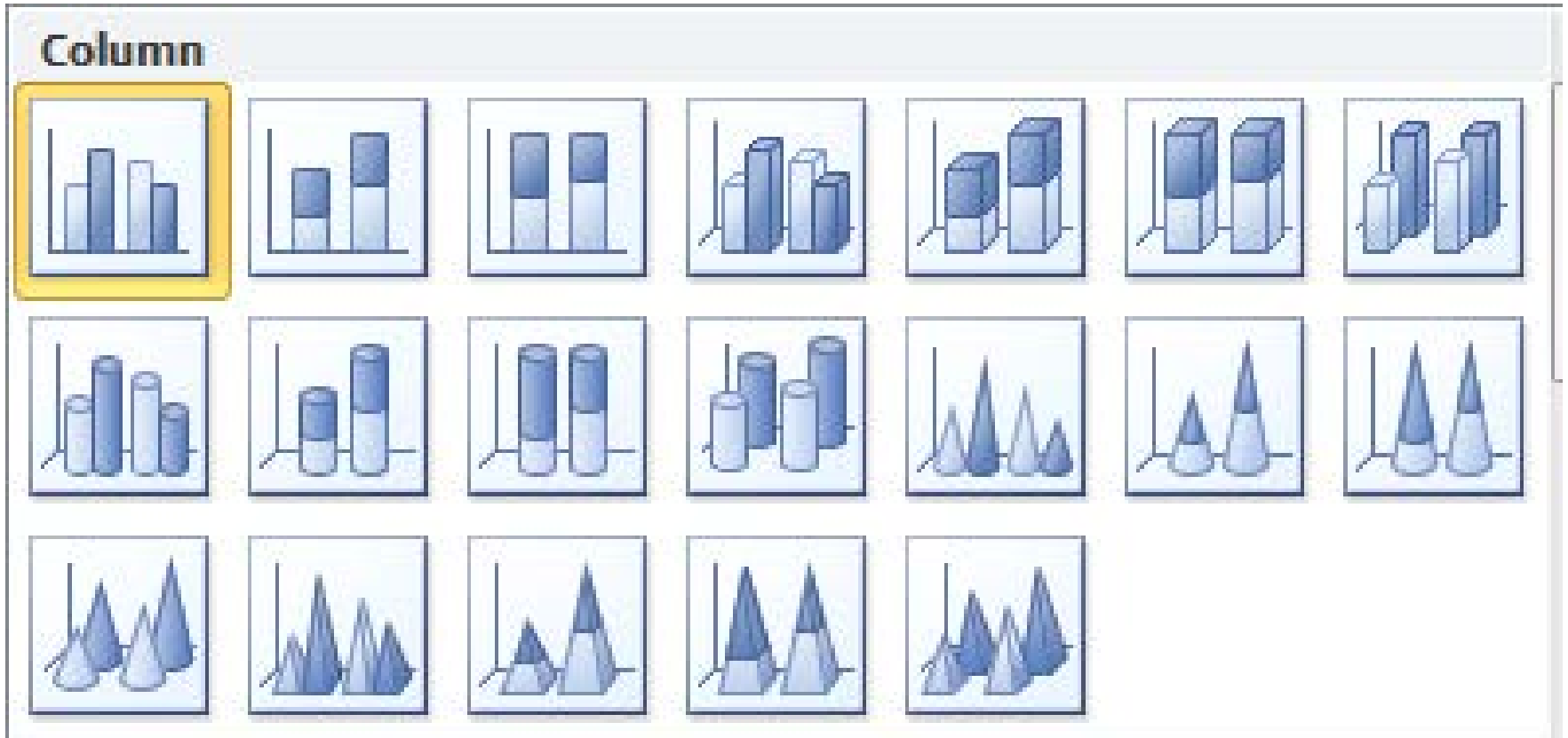
- ❑ Yukarıdaki tabloda yüzdelер verilseydi?
 - ❑ %58'in anlamı ???
 - ❑ %85'in anlamı ???

Grafik Türleri

- ❑ Çubuk
- ❑ Histogram
- ❑ Çizgi
- ❑ Pasta
- ❑ ...

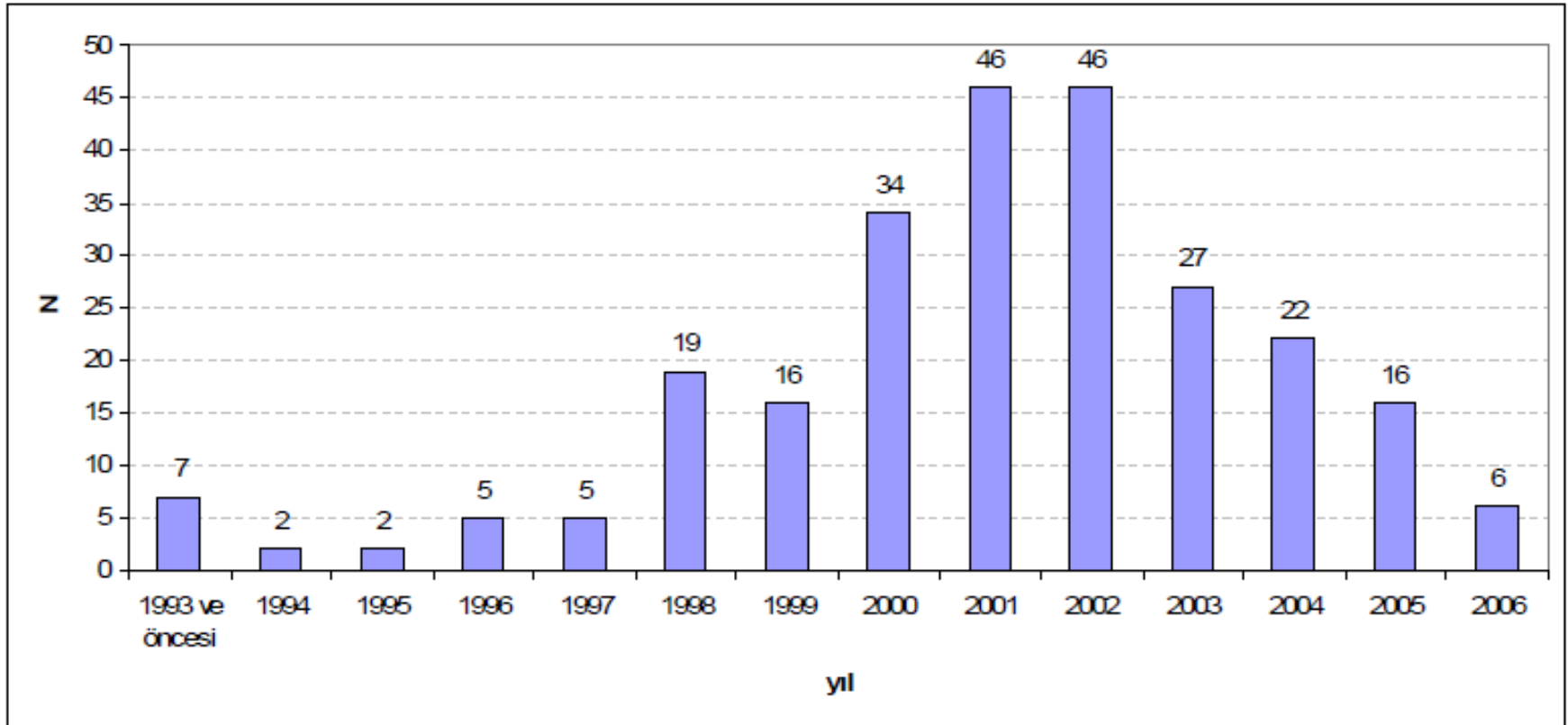
Çubuk Grafik

■ Sıklık ve yüzde



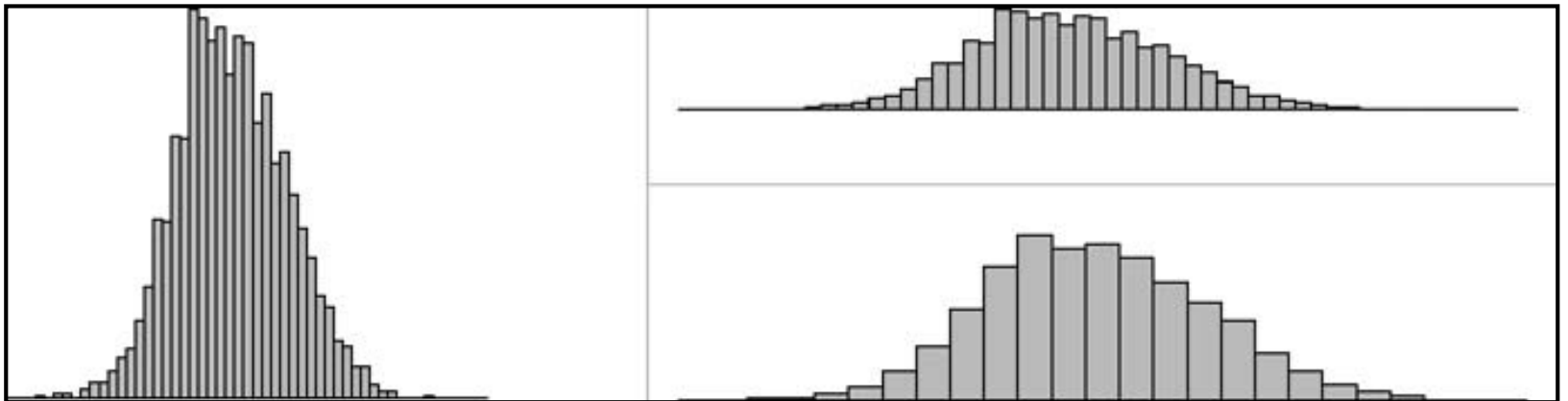
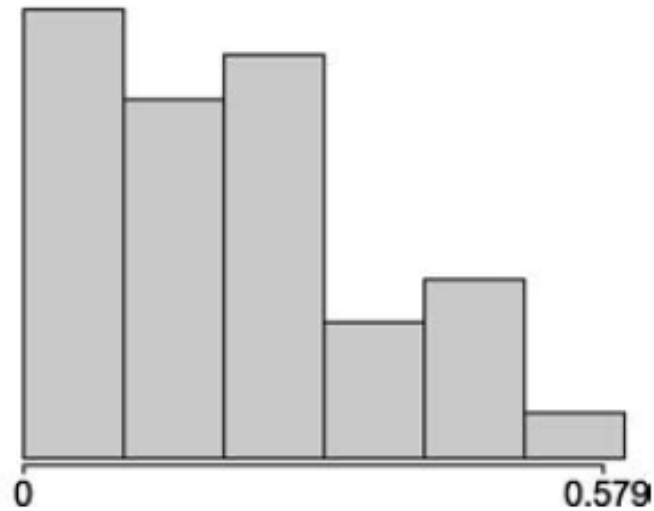
Çubuk Grafik

Dergilerin İlk Kez Elektronik Ortamda Yayımlandığı Yıllar



Histogram

■ Sürekli değişkenler



Histogram

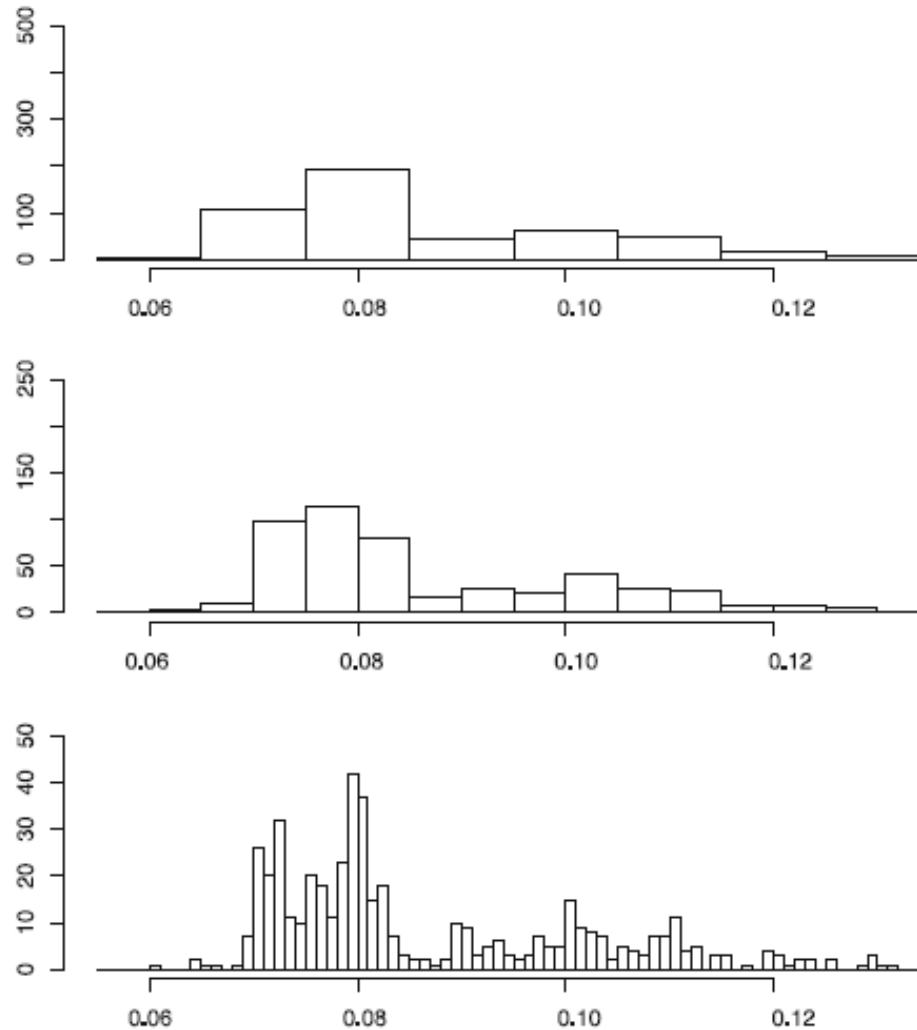
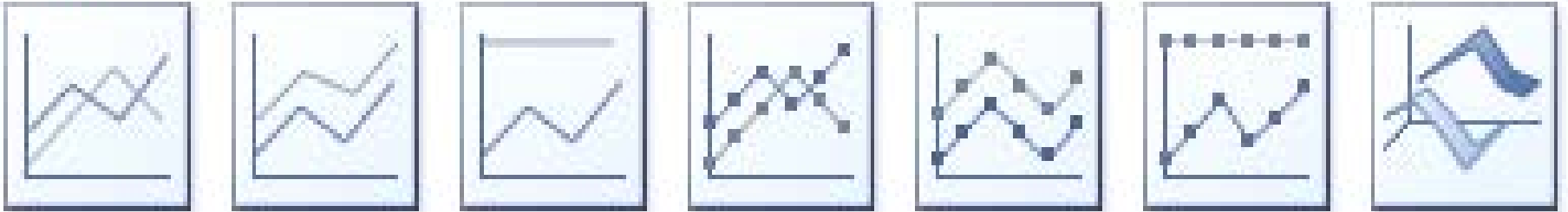


Figure 2.4. Three different histograms of the Hidalgo stamp thickness data, all with the same anchorpoint but with different binwidths. The *horizontal scales* are aligned and the total area of each display is the same (note the different frequency scales). Source: Izenman and Sommer (1988)

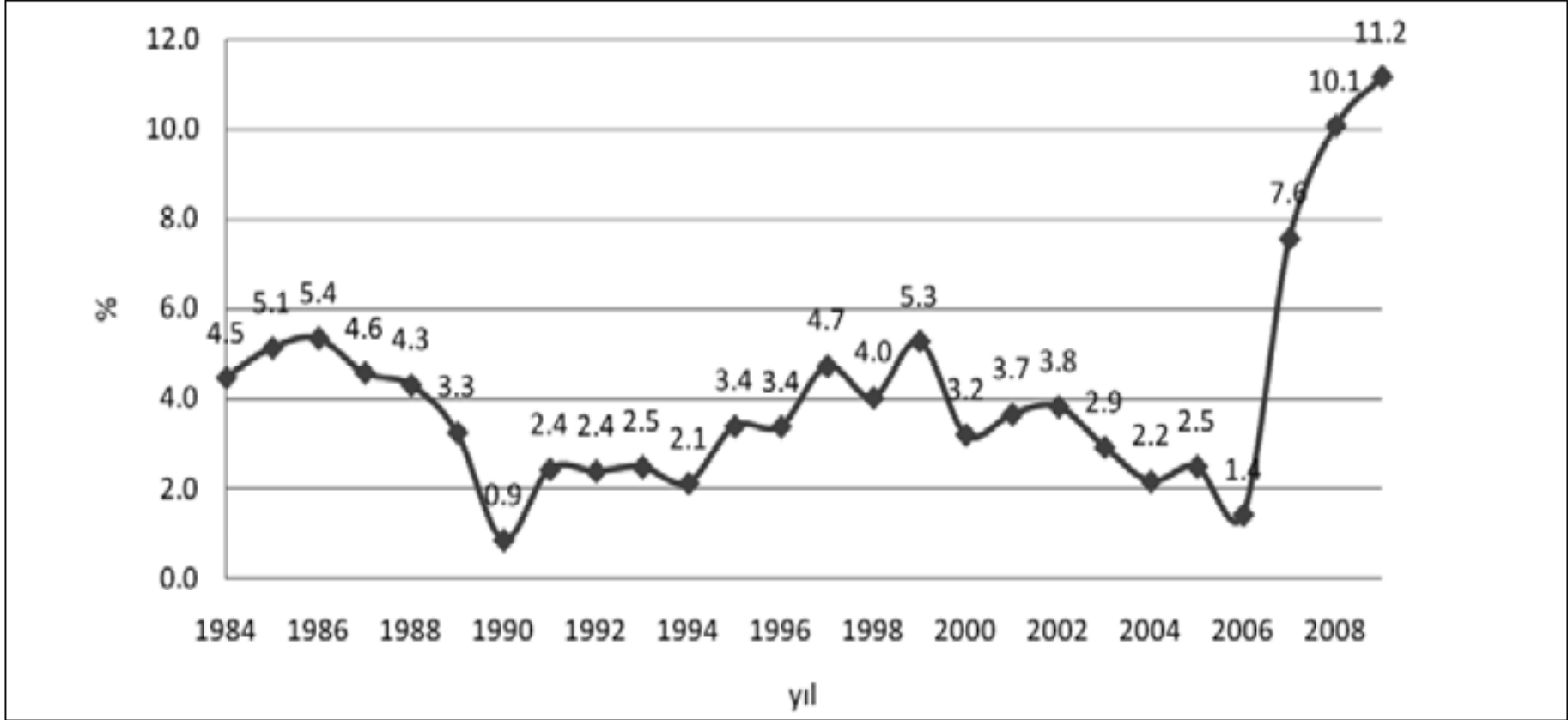
Çizgi Grafik

- ❑ Değişim izlenir
- ❑ Zamana dayalı veriler için uygun

Line



Çizgi Grafik

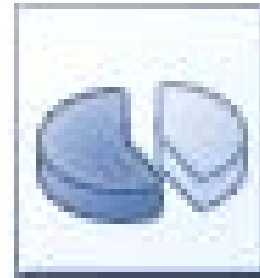
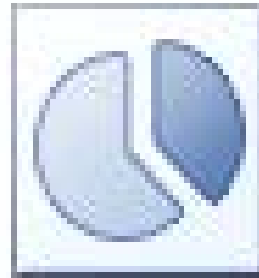


Şekil 2. Türkiye Adresli Dergilerin Türkiye Adresli Yayın Üretimindeki Payı

Pasta Grafik

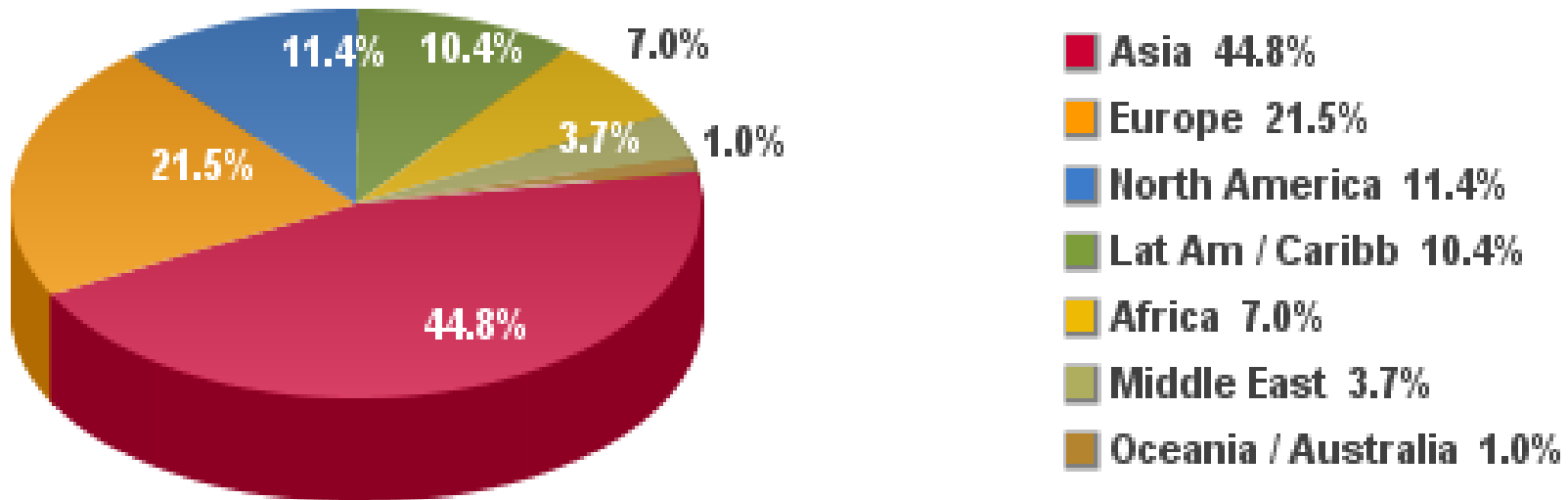
- ❑ Grup verilerine uygun

Pie



Pasta Grafik

Internet Users in the World Distribution by World Regions - 2012 Q2



Source: Internet World Stats - www.internetworldstats.com/stats.htm

Basis: 2,405,518,376 Internet users on June 30, 2012

Copyright © 2012, Miniwatts Marketing Group

Örnekler

When interviewed								
Base			Entering			Leaving		
Gender of respondent			Gender of respondent			Gender of respondent		
Base	Male	Female	Base	Male	Female	Base	Male	Female

Age of respondent									
Base	602	339	263	304	170	134	298	169	129
11-16 years	38	23	15	17	11	6	21	12	9
17-20 years	82	50	32	46	29	17	36	21	15
21-24 years	95	51	44	45	25	20	50	26	24
25-34 years	192	108	84	93	53	40	99	55	44
35-44 years	91	49	42	51	25	26	40	24	16
45-54 years	55	32	23	24	13	11	31	19	12
55-64 years	33	16	17	20	10	10	13	6	7
65+ years	16	10	6	8	4	4	8	6	2

Case	Name	Magnitude	Start	End	Pairs
1	World War I	7.2	1914	1918	44
2	World War II	7.3	1939	1945	86
3	Taiping Rebellion	6.3	1851	1864	1
4	US Civil War	5.8	1861	1865	1
...					
315	Bolivia	2.65	1952	1952	1

Table 7.1 Data on medieval soldiers who served under Henry V during his 1415 expedition into France, at which he fought battles at Harfleur, Calais, and – most famously – Agincourt

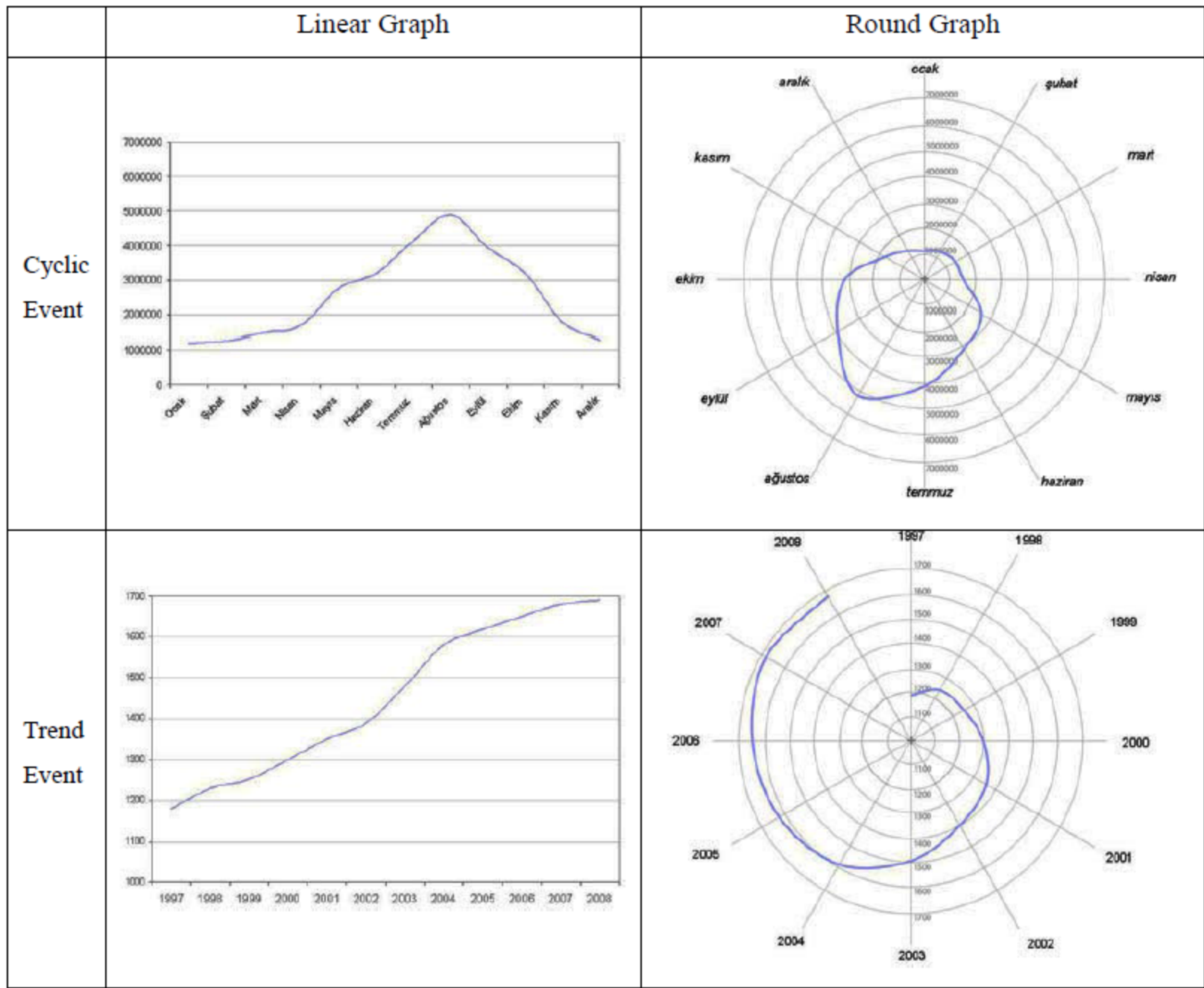
Year	Country	Activity	Rank	Status	Commander	Count
1415	France	Expedition	Man-at-arms	Baron	Henry V	1
1415	France	Expedition	Man-at-arms	Duke	Henry V	3
1415	France	Expedition	Man-at-arms	Earl	Henry V	11
1415	France	Expedition	Man-at-arms	Knight	Henry V	54
1415	France	Expedition	Man-at-arms	Esquire	Henry V	313
1415	France	Expedition	Man-at-arms	Unknown	Henry V	1097
1415	France	Expedition	Unknown	Esquire	Henry V	7
1415	France	Expedition	Unknown	Yeoman	Henry V	15
1415	France	Expedition	Gunner	Unknown	Henry V	2
1415	France	Expedition	Archer foot	Unknown	Henry V	832
1415	France	Expedition	Archer	Yeoman	Henry V	1355
1415	France	Expedition	Archer	Unknown	Henry V	3687

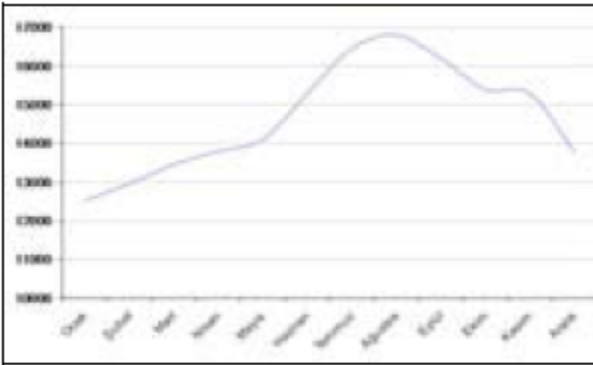
Tablo 12 ULAKBİM kullanıcılarının cinsiyet, yaş, eğitim ve deneyimlerine göre boyutlara bakışları

	Cinsiyet						Yaş						Eğitim						Deneyim														
	Erkek			Kadın			30 ve altı			30 üzeri			Üniv. eğitilmiş			Lisansüstü eğitilmiş			Hiç deneyimi yok			Az deneyimli			Çok deneyimli								
Algı	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N
ULAKBİM	BNG	4,0	1,0	59	4,1	0,9	51	4,0	0,9	36	4,1	0,1	71	3,7	1,2	20	4,2	0,8	88	3,6	1,3	15	4,5	0,4	26	4,0	1,0	69					
	K	3,7	1,1	59	3,6	1,0	51	3,5	1,0	36	3,7	1,1	71	3,5	1,1	20	3,7	1,0	88	3,4	1,2	15	3,8	0,8	26	3,6	1,1	69					
	T	3,2	1,0	59	3,3	0,9	51	3,0	0,9	36	3,4	1,0	71	2,9	0,9	20	3,4	0,9	88	3,1	1,1	15	3,4	0,9	26	3,2	0,9	69					
	Ö	3,5	1,1	59	3,5	1,1	51	3,4	1,0	36	3,6	1,1	71	3,3	1,1	20	3,6	1,0	88	3,1	1,1	15	3,9	0,8	26	3,4	1,1	69					
	Genel $\bar{X}$	3,6	1,1		3,6	1,0		3,5	0,9		3,7	0,7		3,4	1,1		3,7	0,9		3,3	1,2		3,9	0,7		3,6	1,0						
	Beklenti	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS	N	$\bar{X}$	SS
ULAKBİM	BNG	4,6	0,5	59	4,8	0,4	51	4,6	0,6	36	4,7	0,4	71	4,6	0,3	20	4,7	0,5	88	4,5	0,4	15	4,8	0,3	26	4,6	0,5	69					
	K	4,4	0,7	59	4,7	0,5	51	4,5	0,7	36	4,6	0,6	71	4,3	0,8	20	4,6	0,6	88	4,6	0,3	15	4,4	0,8	26	4,6	0,6	69					
	T	3,9	0,8	59	4,0	0,6	51	3,9	0,8	36	3,9	0,7	71	3,6	0,8	20	4,0	0,7	88	3,9	0,7	15	3,6	0,9	26	4,0	0,7	69					
	Ö	4,0	0,8	59	3,9	0,8	51	3,9	0,9	36	4,0	0,7	71	3,9	0,8	20	4,0	0,8	88	3,8	0,5	15	4,1	0,7	26	3,9	0,8	69					
	Genel $\bar{X}$	4,2	0,7		4,4	0,6		4,3	0,7		4,4	0,6		4,1	0,7		4,4	0,7		4,3	0,5		4,2	0,7		4,3	0,7						

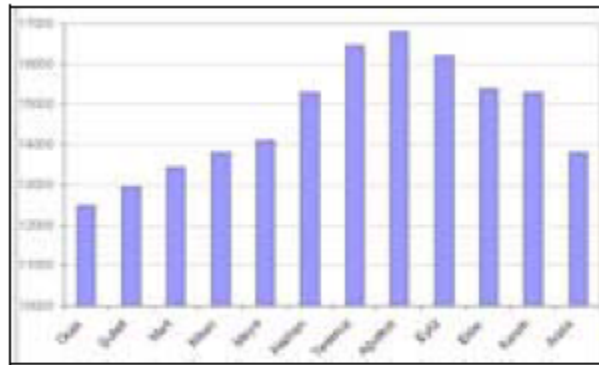
Not:1. Boyut adları BNG (Bilginin niteliği ve güven), K (Kullanılabilirlik), T (Tasarım), Ö (Özdeşleştirme) şeklinde kısaltılmıştır.

2. ULAKBİM kullanıcıları arasında üniversite ve lisansüstü eğitim almamış olan yalnızca üç katılımcı vardır (Bkz. Tablo 10b). Bu nedenle ULAKBİM kullanıcılarının eğitim durumları ile ilgili karşılaştırmalar yalnızca üniversite ile Lisansüstü eğitilmişler arasında yapılmıştır.

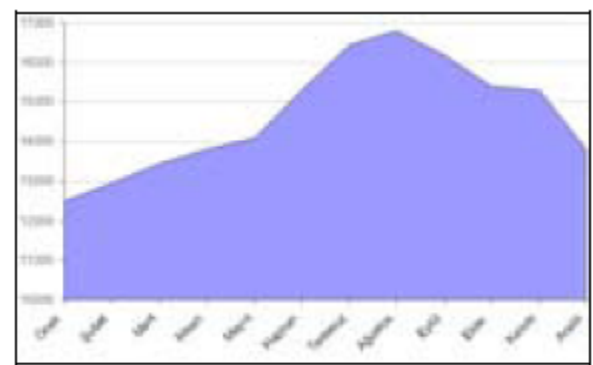




Çizgi Grafik



Sütun Grafik



Alan Grafik

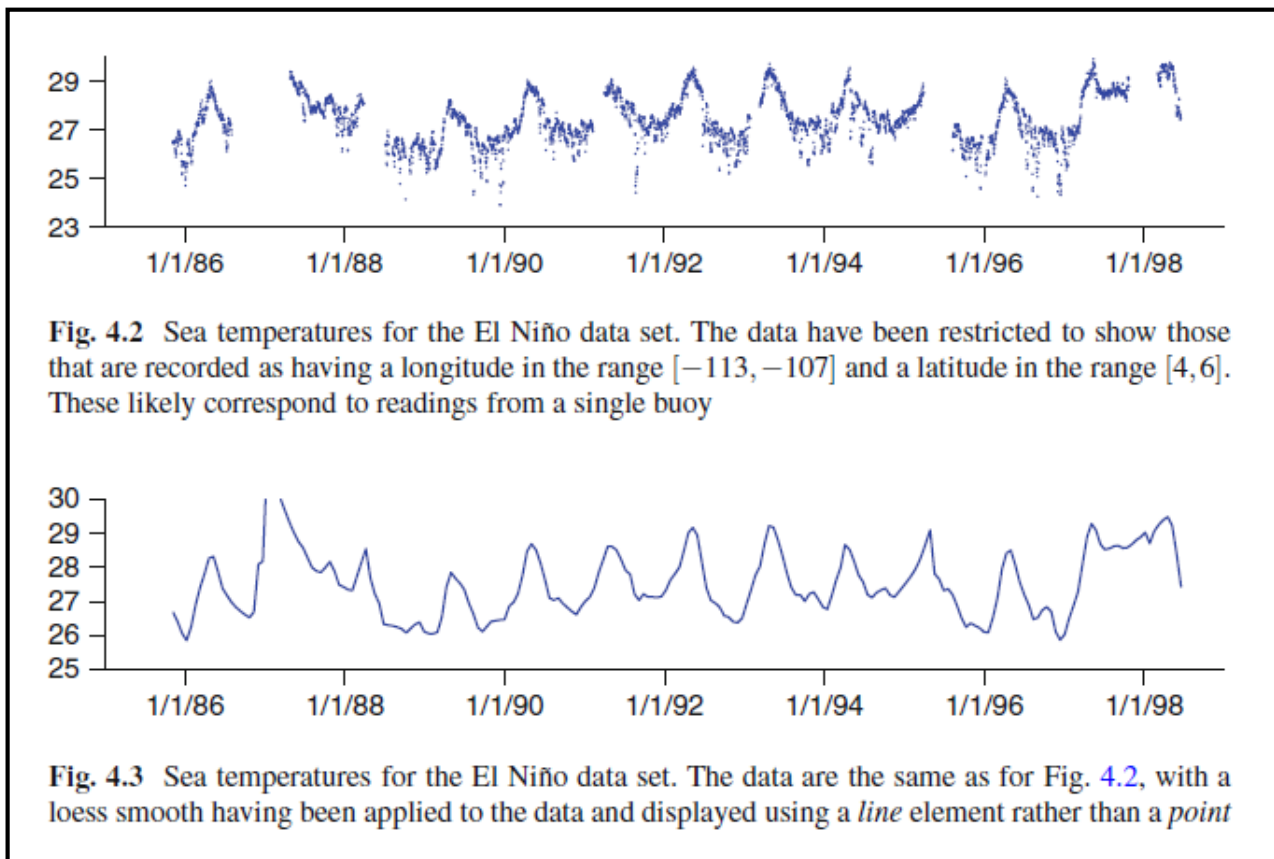


Fig. 4.2 Sea temperatures for the El Niño data set. The data have been restricted to show those that are recorded as having a longitude in the range $[-113, -107]$ and a latitude in the range $[4, 6]$. These likely correspond to readings from a single buoy

Fig. 4.3 Sea temperatures for the El Niño data set. The data are the same as for Fig. 4.2, with a loess smooth having been applied to the data and displayed using a *line* element rather than a *point*

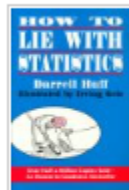
How to Lie with Statistics



By
DARRELL HUFF

Pictures by IRVING GELS

1. How to lie with statistics / by Darrell Huff ; pictures by Irving Geis.



Kitap

By: Huff, Darrell. New York : W.W. Norton, 1993. 01/01/1993 142 s. Language: English, Veritabanı: Hacettepe Üniversitesi Kütüphaneleri Kataloğu

Konular: İstatistik; Statistics

Konum	Yer No.	Durum	Son Tarih
Beytepe Genel Koleksiyon	HA29 .H82 1993	Rafta	
Daha Fazla Göster (1)			

RASMUSSEN REPORTS POLL

Did scientists falsify research to support their own theories on Global Warming?

59%

SOMEWHAT LIKELY

35%

VERY LIKELY

26%

NOT VERY LIKELY



LIMATE CHANGE RESEARCH / FOX NEWS\ GOP S NHL TOR 6 COB 3

2012 PRESIDENTIAL RUN

GOP CANDIDATES



FOX

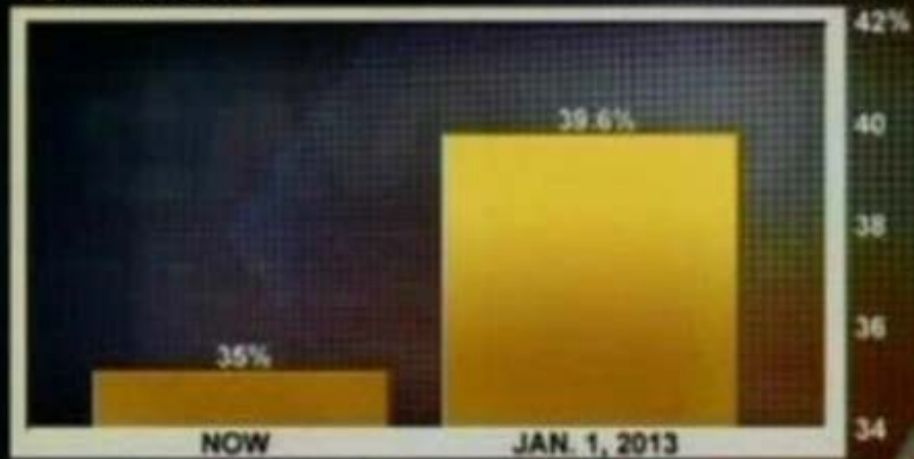
47'

SOURCE: OPINIONS

DYNAMIC

IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

CONSUMER

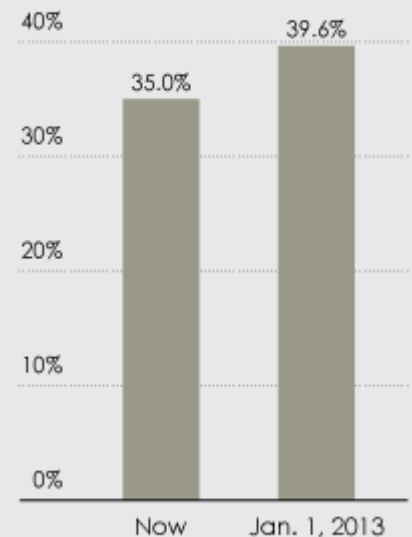
WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

DOW 13008.68 ↑ 64.33 S&P 1379.32 ↑ 5.98 NASDAQ 2939.52 ↑ 6.32

Woah, that's a big difference!

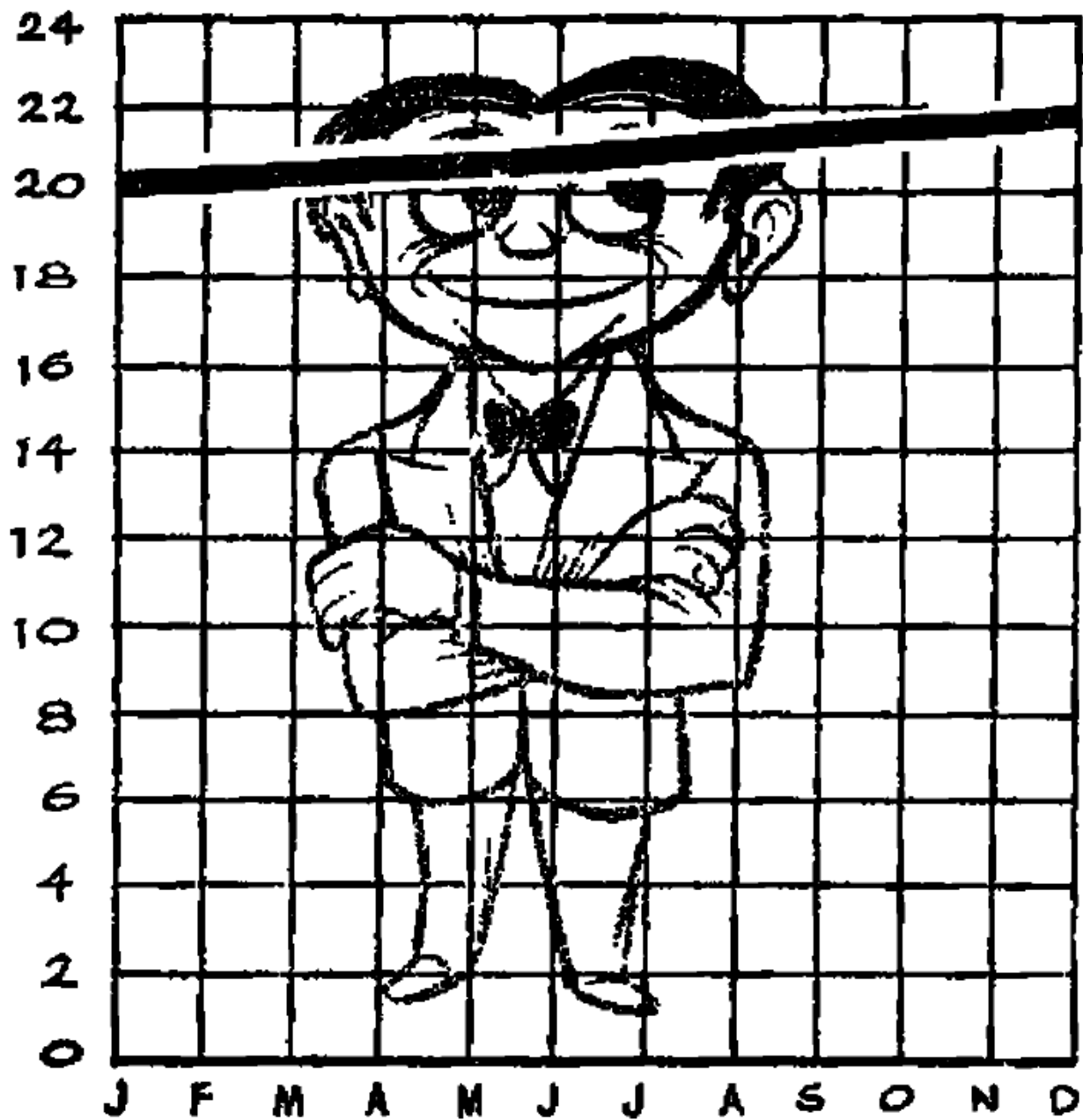
If Bush tax cuts expire...

Top tax rate

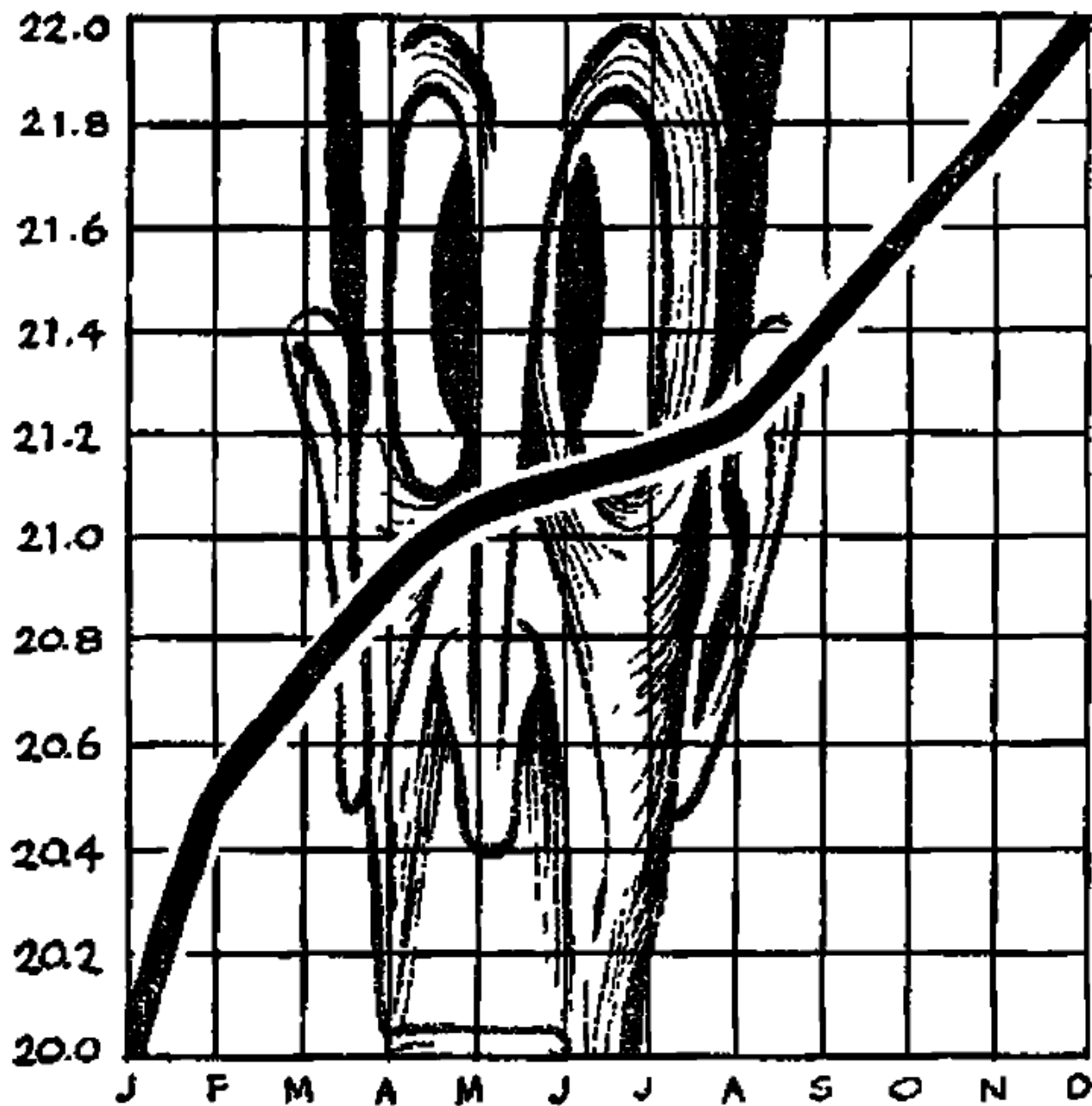


Now, that's more like it...

Billion Dollars



Billion Dollars



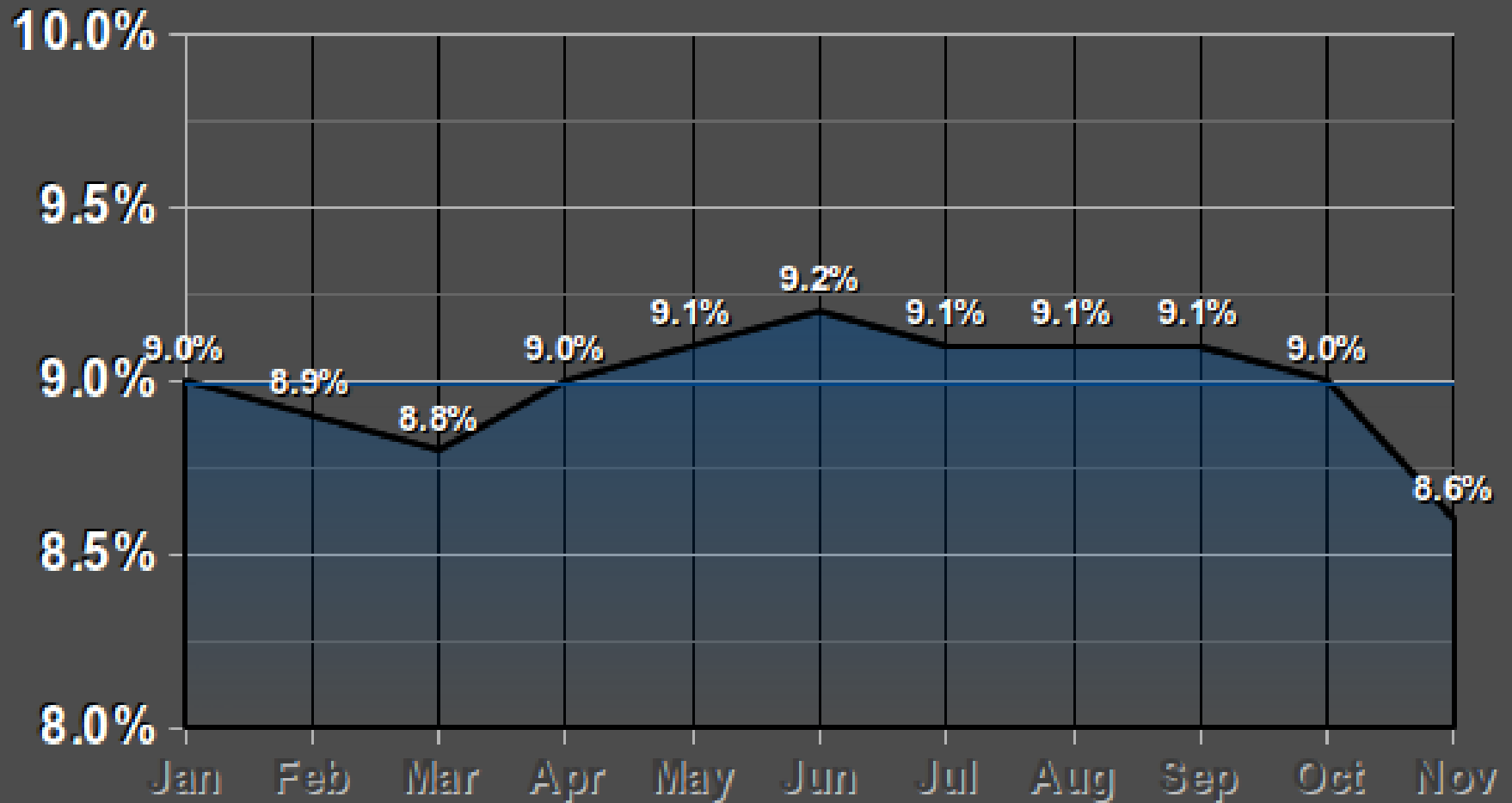
UNEMPLOYMENT RATE UNDER PRESIDENT OBAMA



2011

SOURCE: BUREAU OF LABOR STATISTICS

AR WAR AND A TROOP WITHDRAWAL AT THE E NAS FUT 2,292.50



JOB LOSS BY QUARTER



FOX NEWS
FOX NEWS
.COM

SOURCE: BLS

AMERICA'S
NEWSROOM

UNEMPLOYMENT LEVEL BY RANDOM QUARTER



**50%
PAY CUT**

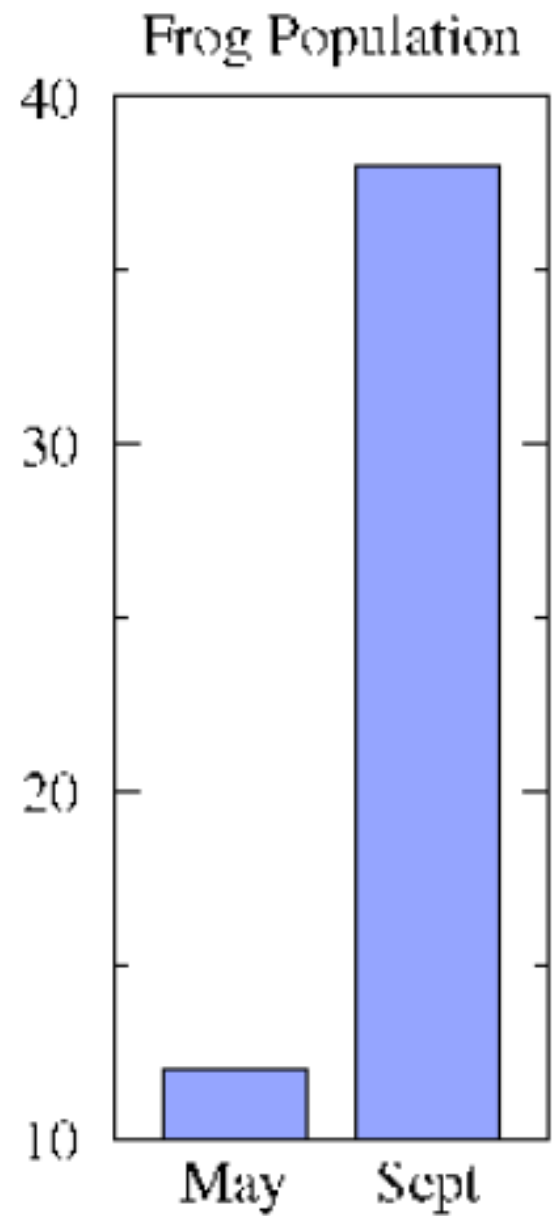
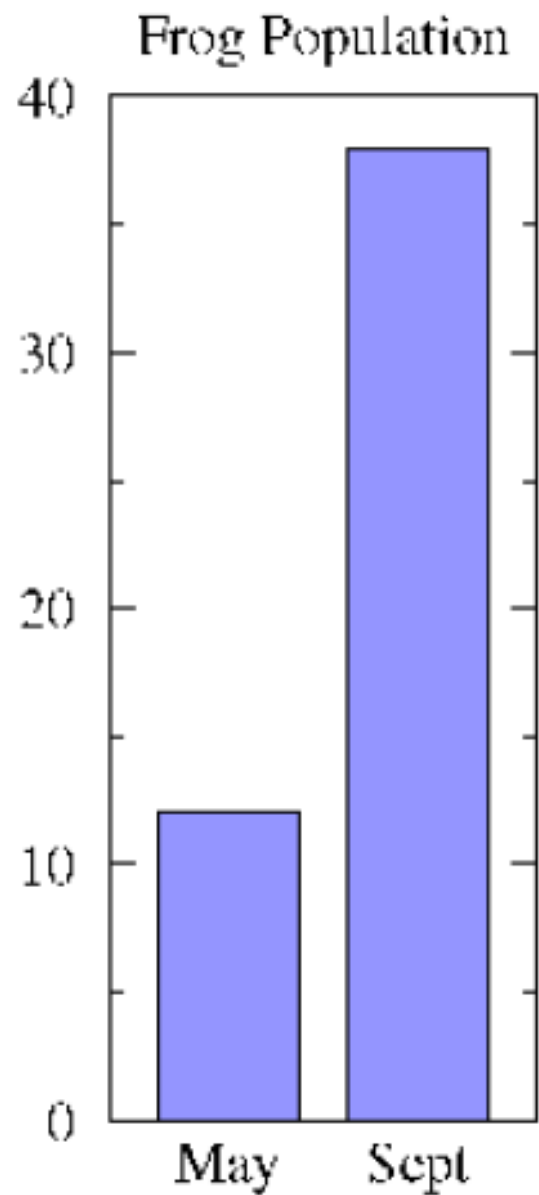
$\begin{array}{r} \$100 \\ 50\% \\ \hline \$50 \end{array}$	$\begin{array}{r} \$100 \\ -50 \\ \hline \$50 \end{array}$
---	--

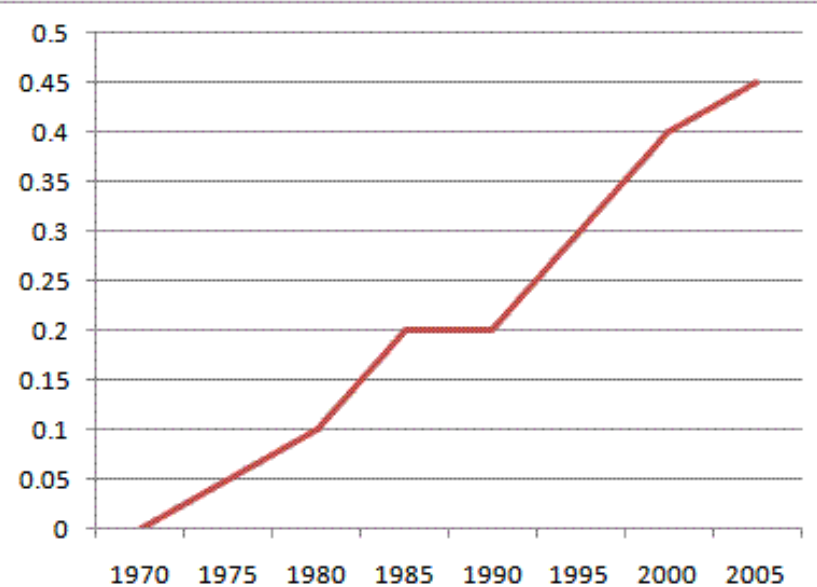
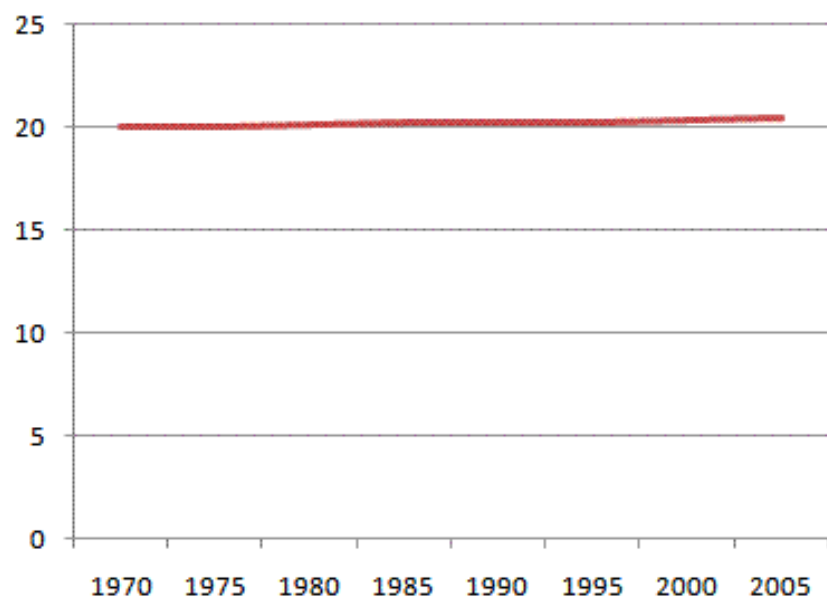
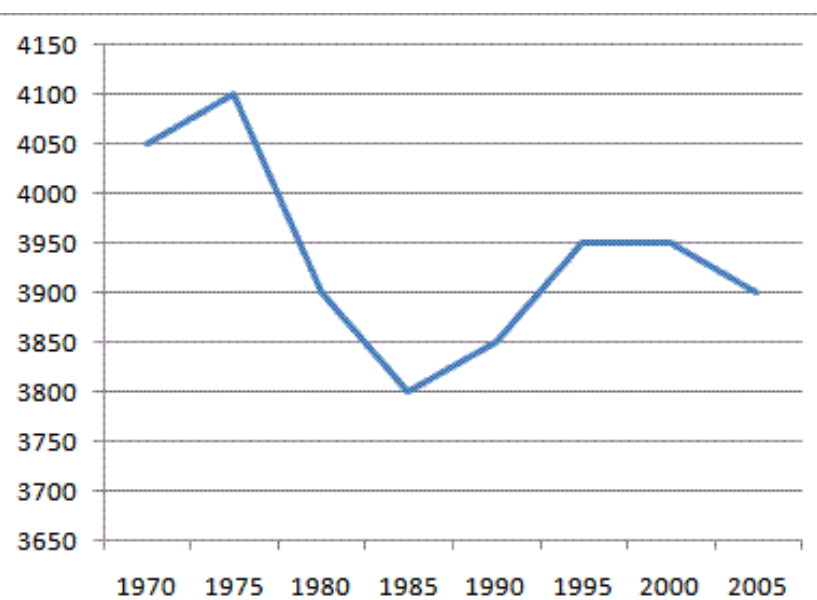
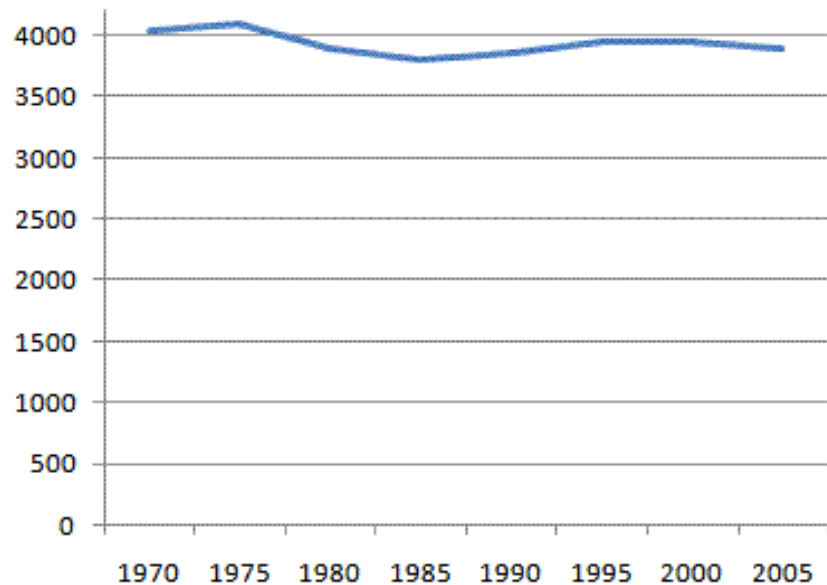


**50%
PAY CUT
RESTORED**

$$\begin{array}{r} \$50 \\ 50\% \\ \hline \$25 \end{array} \qquad \begin{array}{r} \$50 \\ 25 \\ \hline \$75 \end{array}$$







There's a misrepresentation for that

Verizon's ads suggest that AT&T has no data coverage in much of the country by comparing only AT&T's newer, faster 3.2 Mbit 3G network against its own 1.4 Mbit 3G EVDO while excluding any mention of AT&T's slower but functional EDGE service, as well as its free WiFi access points.



Verizon's advertised 3G maps



AT&T's own data coverage map

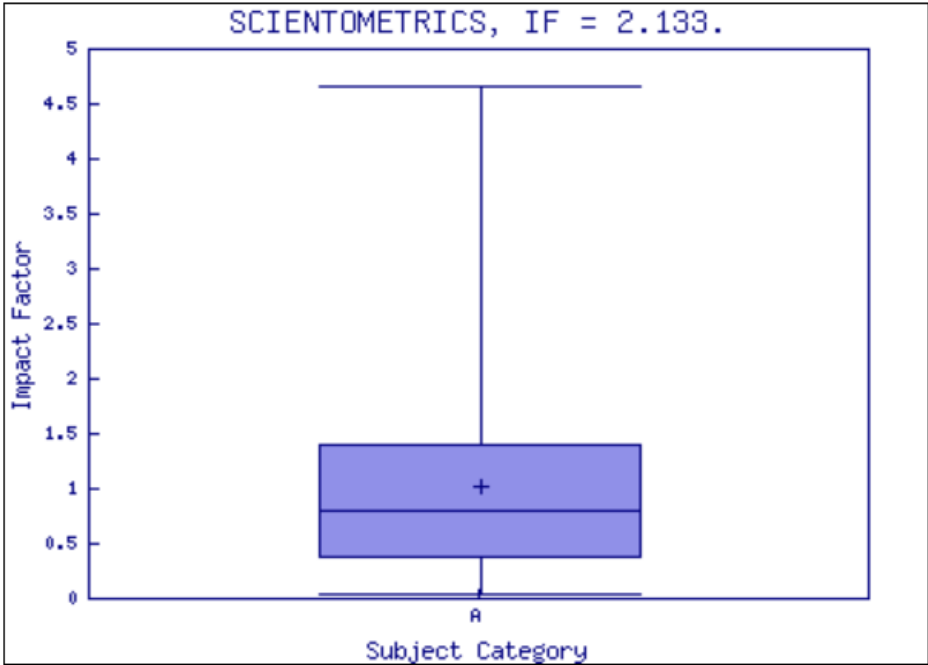
A comparison of US population density explains why AT&T hasn't focused its 3G deployment on Nevada and rural areas of the Mountain Time Zone.



Actual US population map

So who really has more coverage?

This is a box plot of the subject category or categories to which the journal has been assigned. It provides information about the distribution of journals based on Impact Factor values. It shows median, 25th and 75th percentiles, and the extreme values of the distribution.



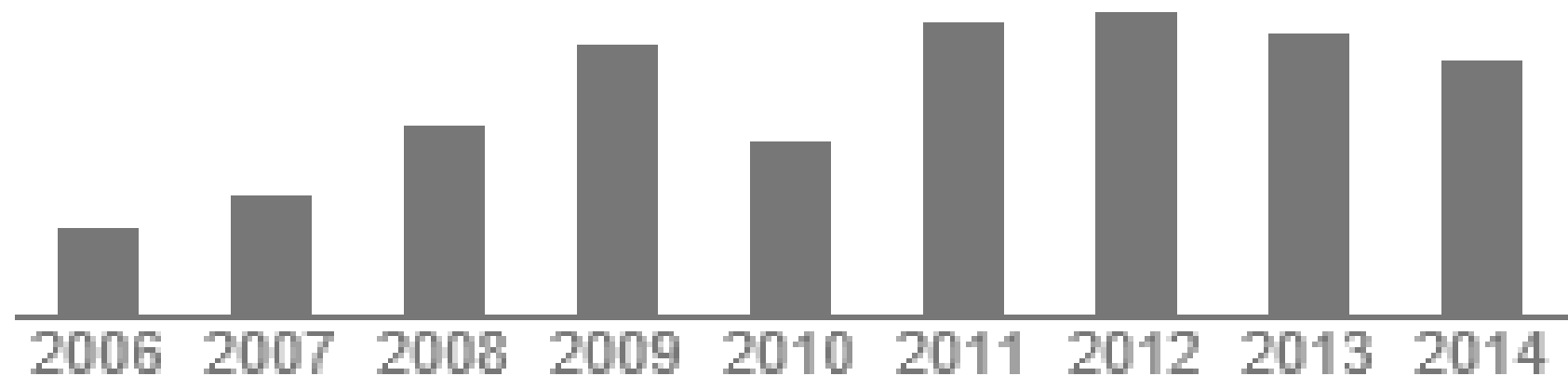
Key
A - INFORMATION SCIENCE & LIBRARY SCIENCE

This table shows the ranking of this journal in its subject categories based on Impact Factor.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
INFORMATION SCIENCE & LIBRARY SCIENCE	85	7	Q1

Citation indices

	All	Since 2009
Citations	411	304
h-index	12	9
i10-index	16	9



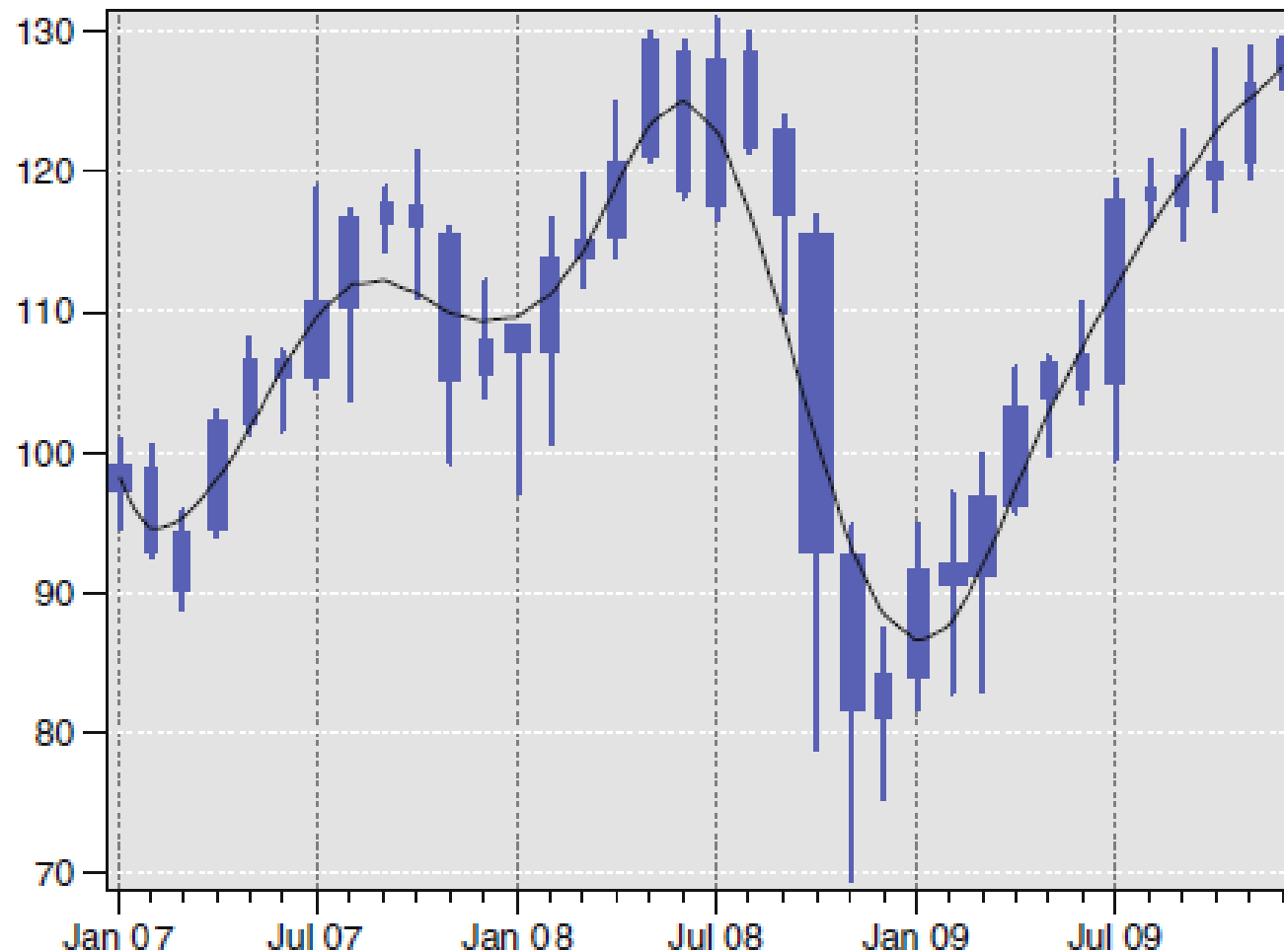


Fig. 8.7 IBM stock. This data set contains monthly data on IBM stock for several years. One edge element links the high and low prices for the month (*thin line*). Another edge element links the open and close prices, shown as a *thicker line*. The width of that *thicker line* denotes the volume. A *smooth line* is an explicit estimate for the actual price, smoothing the start and end prices for the data

Güncü

10 Gün

1 Ay

3 Ay

1 Yıl

Bu Yıl

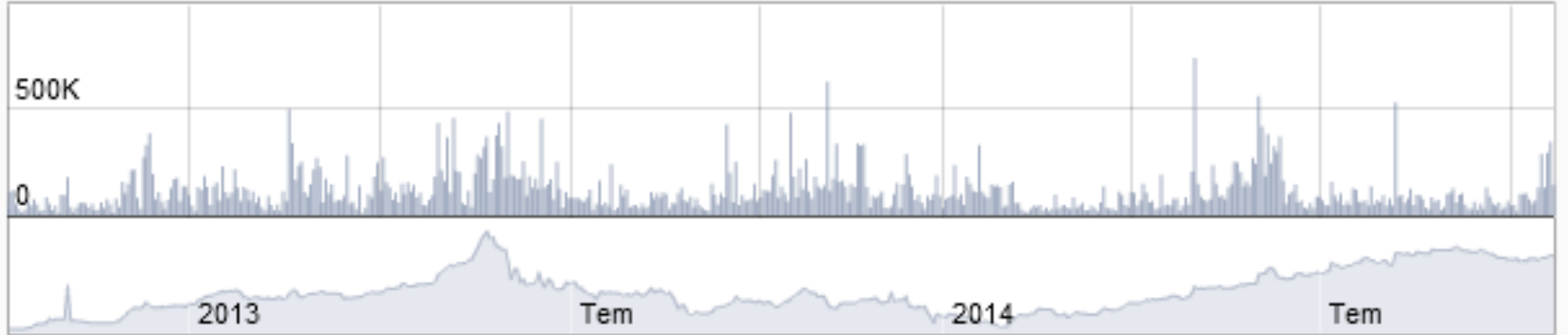
Tümü

Değer

Eki 03, 2012 - Eki 23, 2014



Hacim En Düşük: 7 398 En Yüksek: 742 167



Periyod: TUMU

Güncü

10 Gün

1 Ay

3 Ay

1 Yıl

Bu Yıl

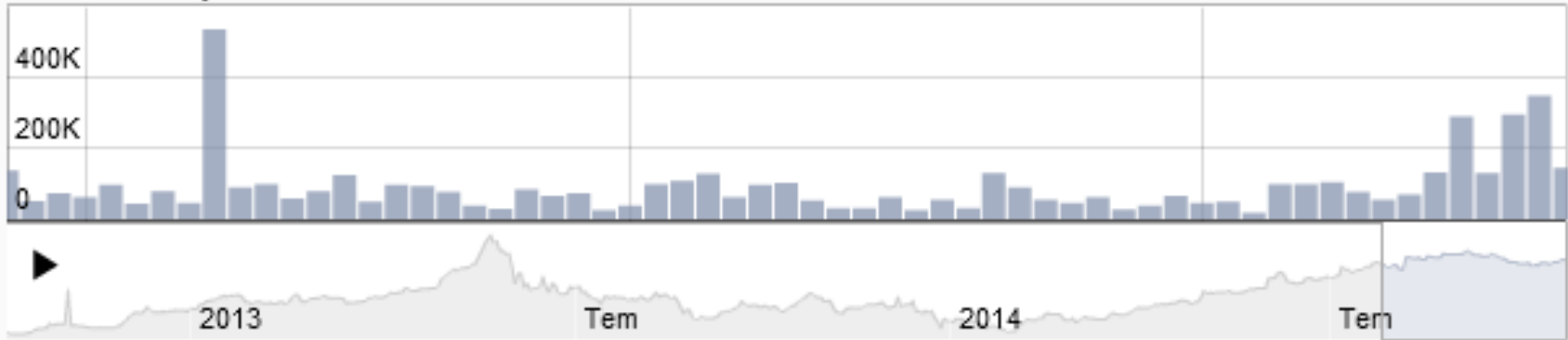
Tümü

Değer

Tem 24, 2014 - Eki 23, 2014



Hacim En Düşük: 18 473 En Yüksek: 532 383



Periyod: 3 AY

Güncü

10 Gün

1 Ay

3 Ay

1 Yıl

Bu Yıl

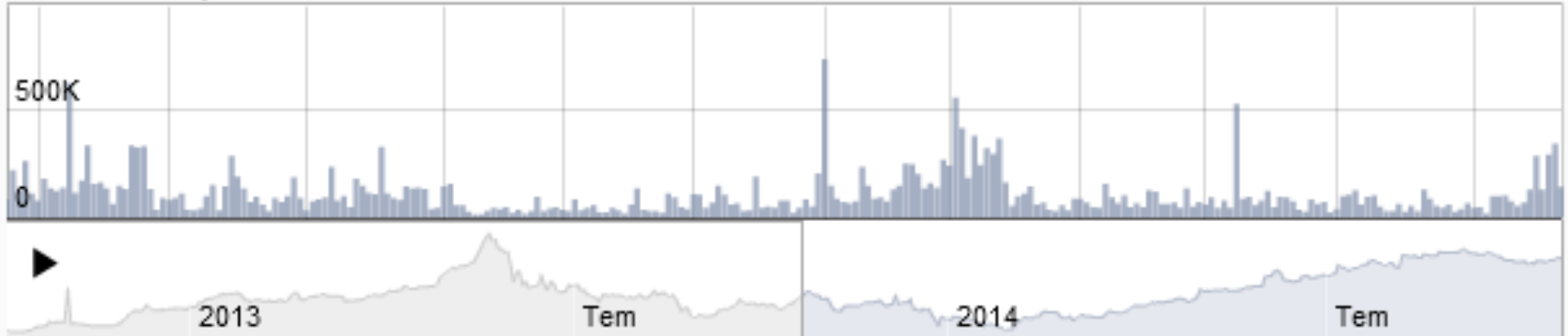
Tümü

Değer

Eki 24, 2013 - Eki 23, 2014

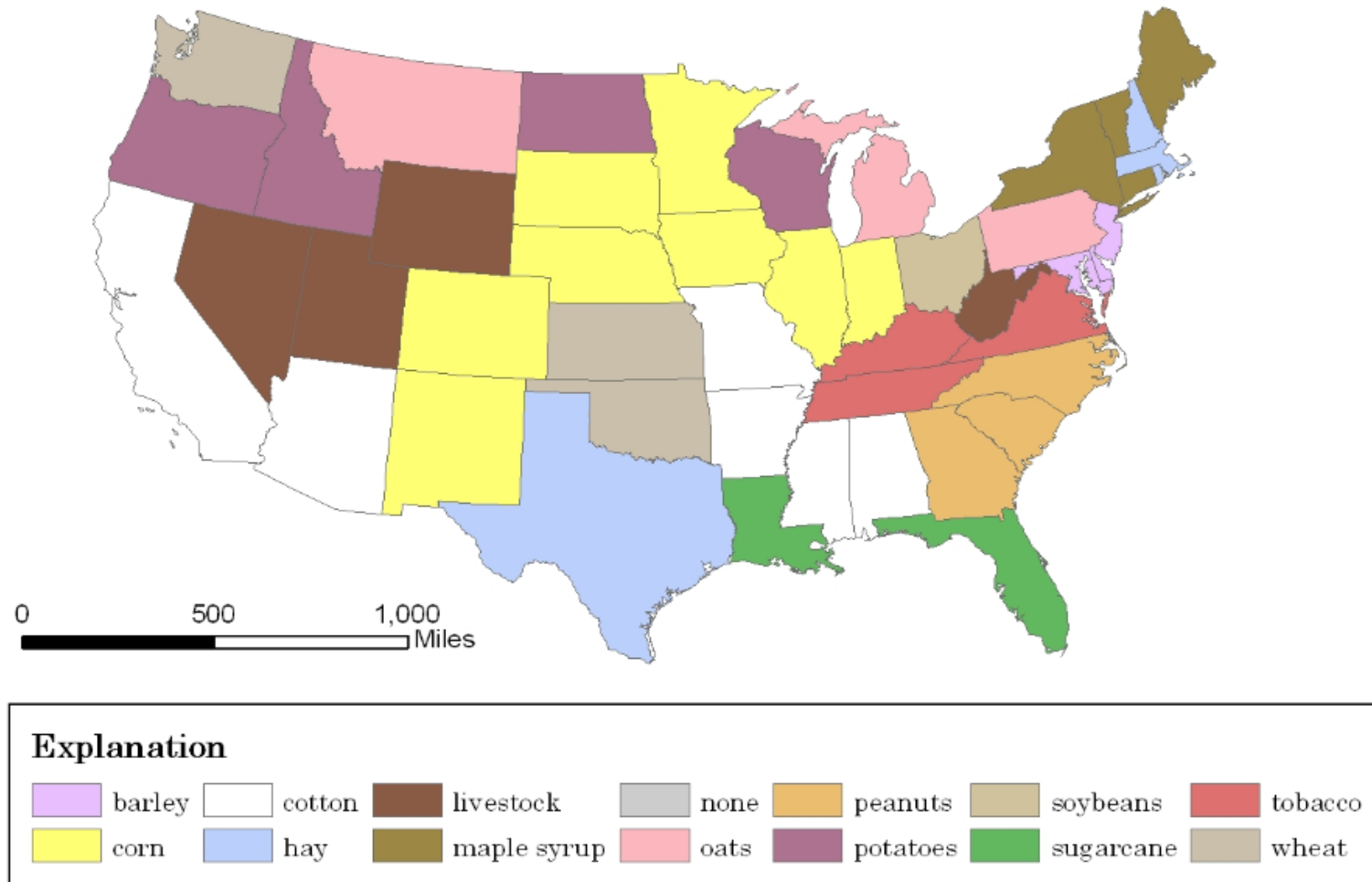


Hacim En Düşük: 15 651 En Yüksek: 742 167



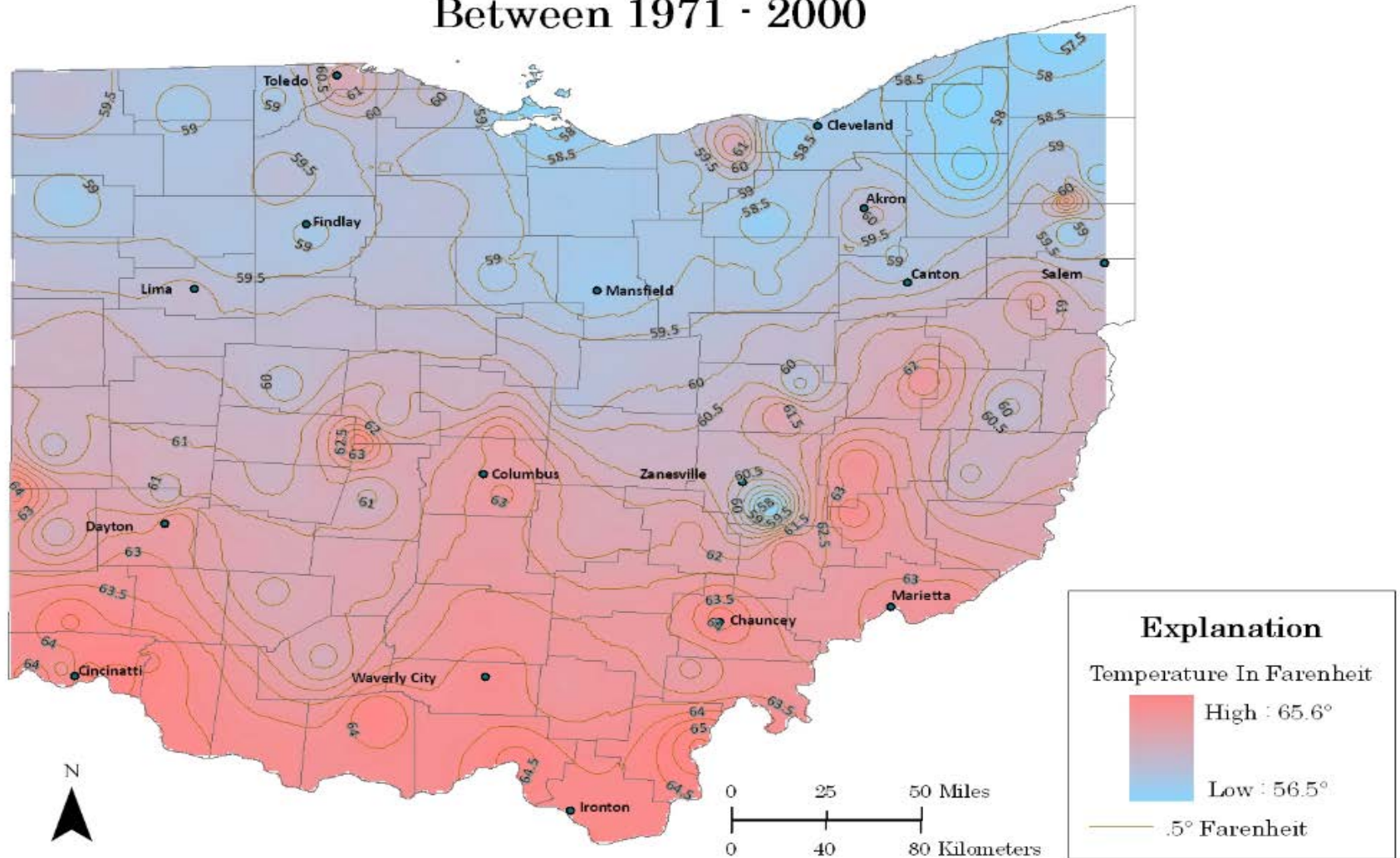
Periyod: 1 YIL

United States Farms Agriculture Production 2007



Map Created By: Victoria Kenton 10/13/09
 Source: <http://nass.usda.gov/QuickStats/>

Average Temperature in Ohio Between 1971 - 2000



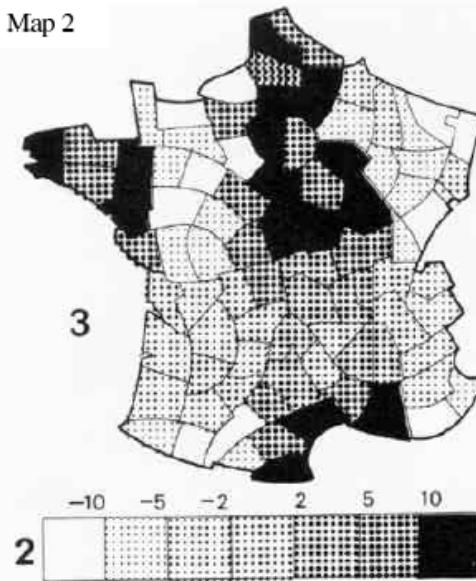
Map Created By: Victoria Kenton 11/11/09
Source:
NOAA's National Climatic Data Center

Rainbow Scale Considerations

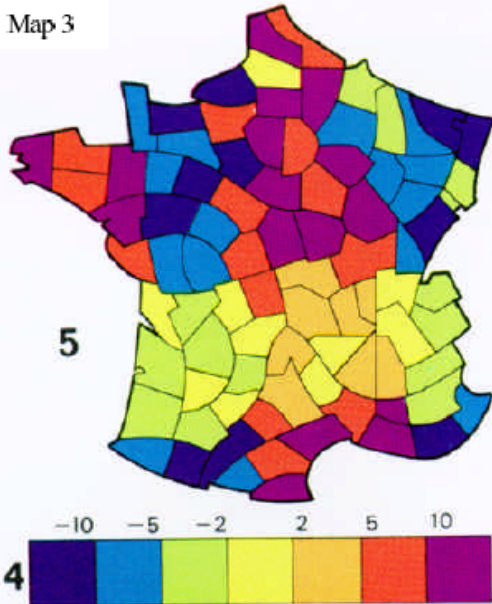
Map 1



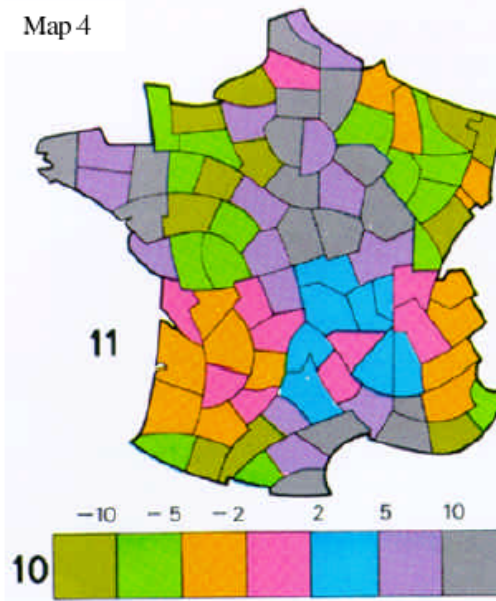
Map 2



Map 3



Map 4



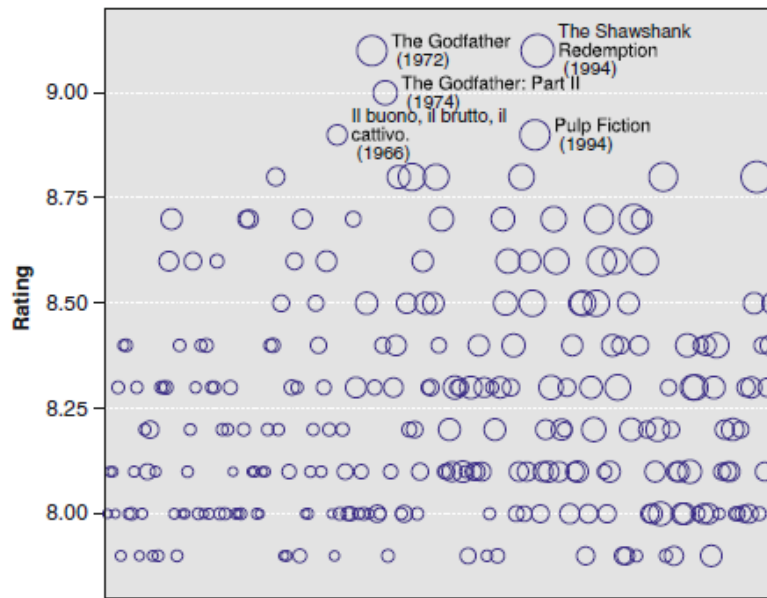


Fig. 8.2 Top 250 movies as of January 2010. This chart shows the highest rated movies as a scatterplot of ratings by time. The number of votes for each movie has been mapped to the size of the points. In this version of the figure the dates are shown as an ordered list, with the x dimension simply showing the order within the data set

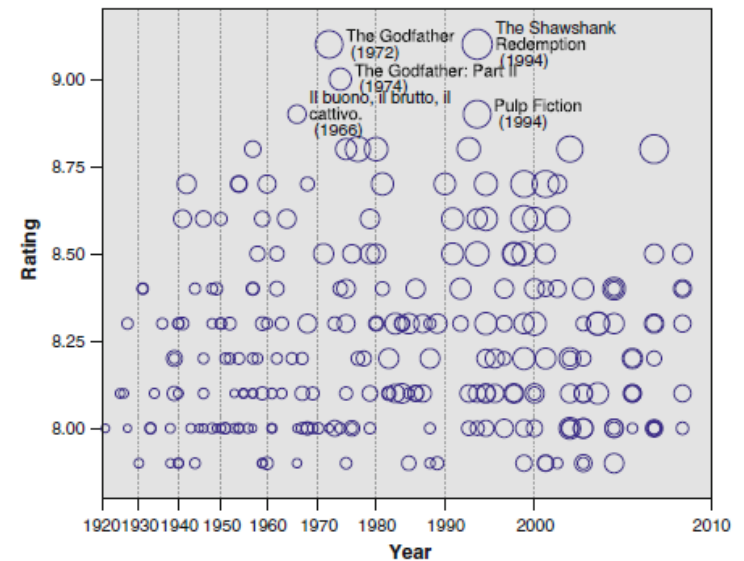


Fig. 8.3 Top 250 movies as of January 2010. This chart shows the highest rated movies as a scatterplot of ratings by time. The number of votes for each movie has been mapped to the size of the points. A square root transform has been applied to both the x dimension and the size aesthetic. For the size, this is so that the area represents the counts. For the dimension the purpose is different; the x dimension has been transformed so that more recent movies are given more space to be displayed – the time scale has been distorted to focus attention on the important time period

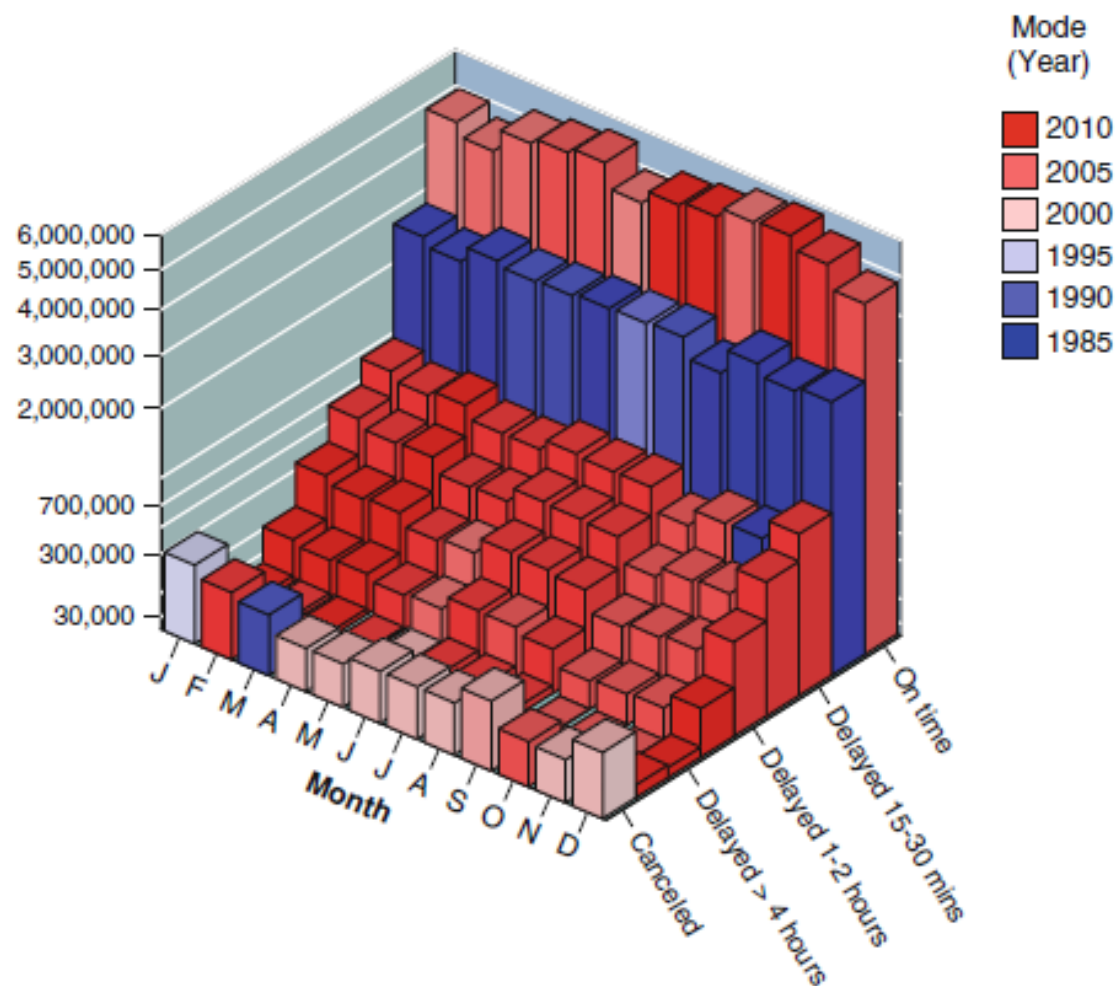
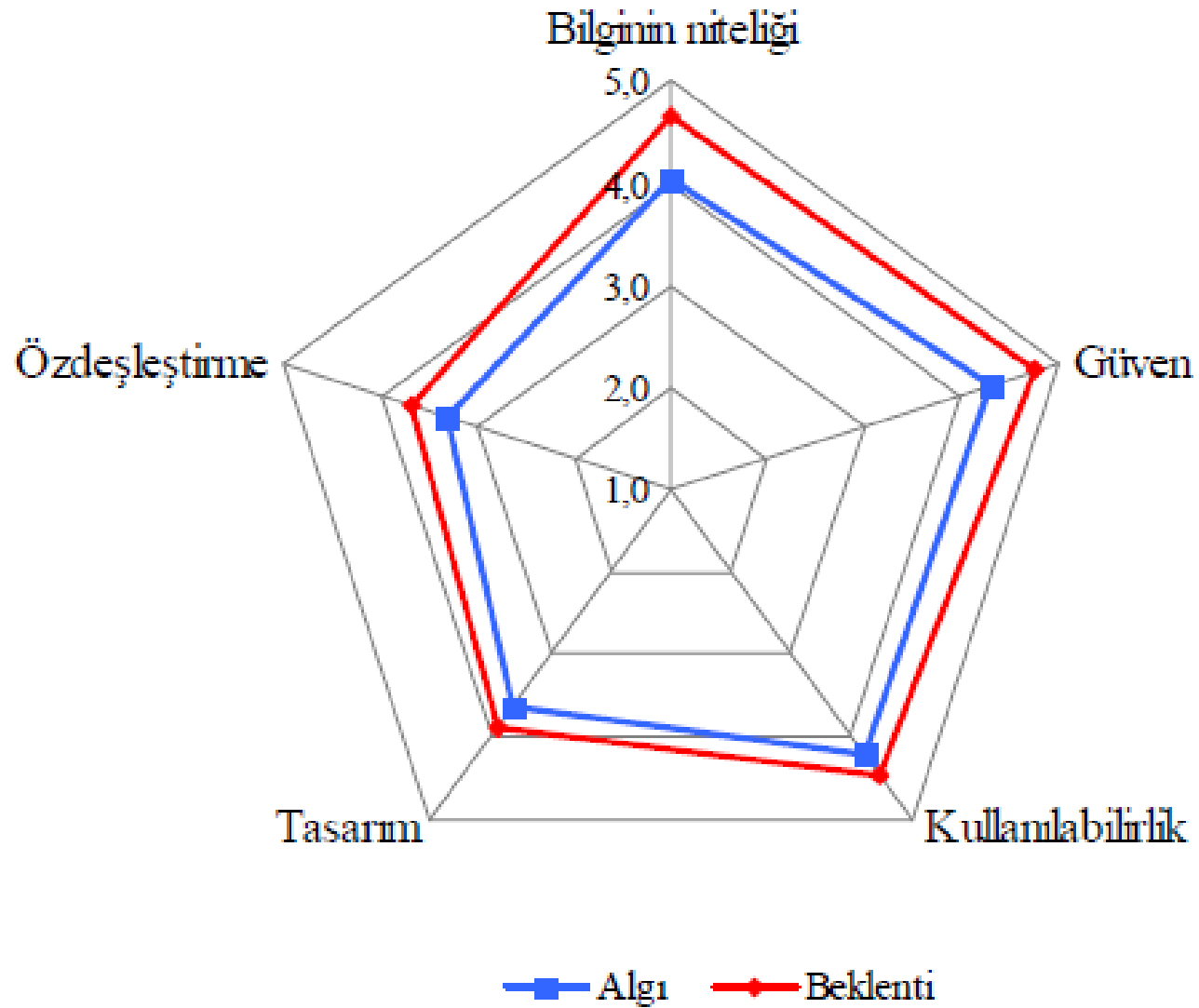


Fig. 10.1 Delay counts by severity and month of year. The data for this figure were rolled up in the database. SQL versions of the date transformations given in Table 8.1 were used to calculate the month in which each flight took place, and the delay duration was categorized into one of nine different ranges. For each resulting table cell, the most commonly occurring year was also recorded and used to color the 3-D bar chart. Note also the nonlinear scale on the *vertical axis*, allowing us to see features in the less-frequent delay classifications. The year has been encoded as color on a divergent scale, with early years in *blue*, later years in *red*, and years in the middle in *white*



Şekil 5 Idefix için belirlenen boyutların algı ve beklenti puanlarına göre karşılaştırılması

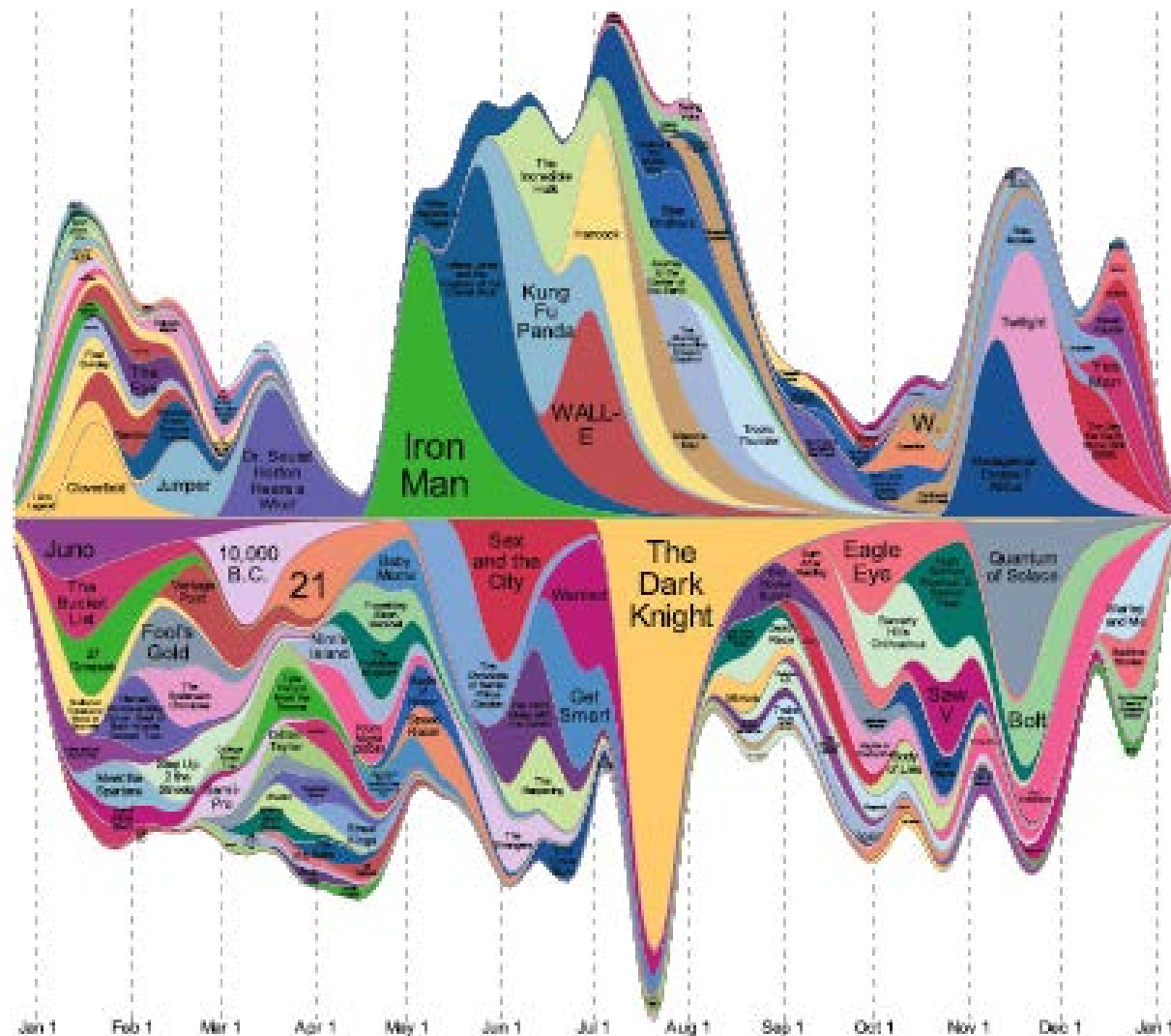


Fig. 5.9 US box-office receipts for movies in 2008. Time runs horizontally in this 1-D chart. Each movie is represented by a density estimate for when people saw the movie, and the densities are stacked on top of each other; the color used for each movie has no meaning; it is simply used to distinguish different movies. This figure is very compact due to space constraints; when stretched out four to eight times as wide, the riverlike nature of the chart is more apparent

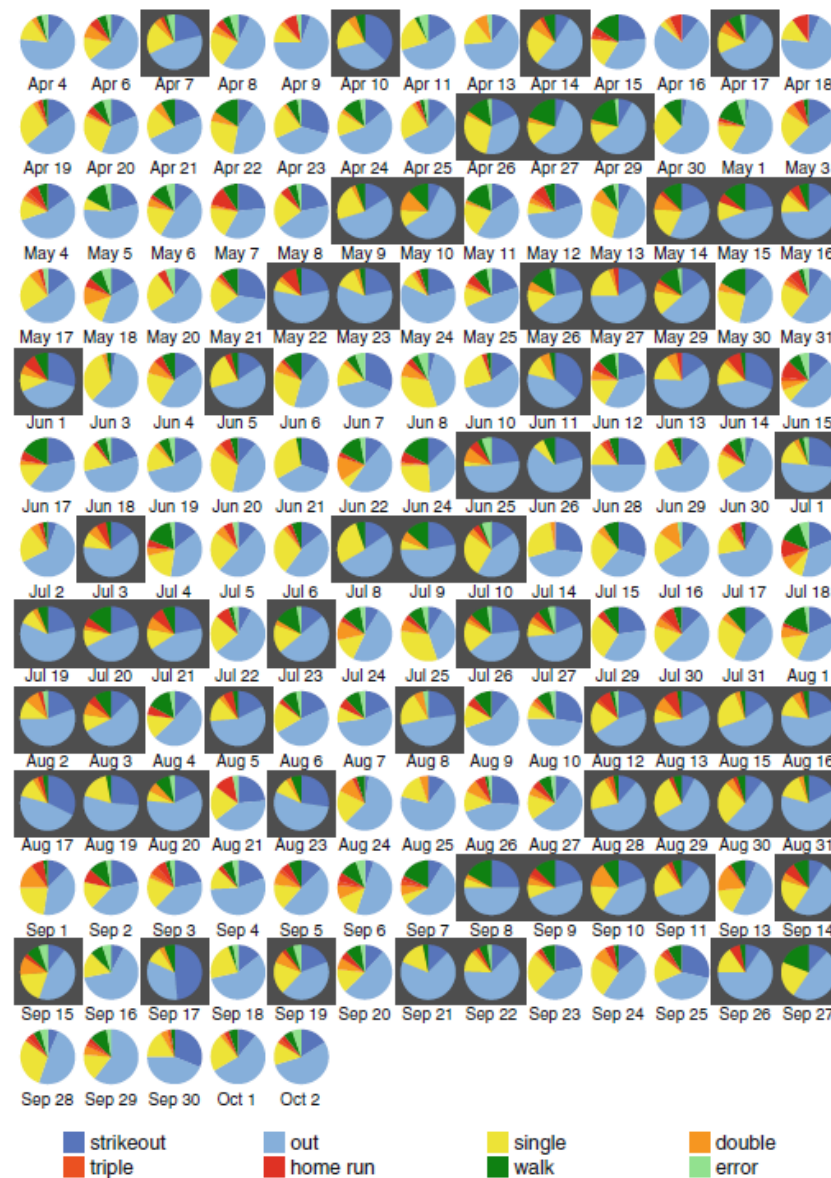


Fig. 6.14 White Sox plays made in the 2005 season. This figure shows pie charts indicating the frequencies of different types of play, with one pie chart for every game in the season. This was achieved by faceting the basic chart by the *date* variable. A *brightness aesthetic* has been applied to the cells that darkens the cell background for games that were lost

Tablo ve Grafik Yaparken Dikkat Edilmesi Gerekenler

- ❑ Tablo numarası ve tablo başlığı
- ❑ Tablo başlıkları tablonun üstünde, grafik başlıkları grafiğin altında yer almalı
- ❑ Tablo başlığı tablonun içeriğini tanımlamalı ve kısa olmalı
- ❑ Tabloda bağımsız değişken satırda, bağımlı değişken sütunda yer almalı
- ❑ Grafiklerde X eksenini ve Y eksenine uygun veriler konmalı => X eksenini yıl, Y eksenini sıklık gibi

Tablo ve Grafik Yaparken Dikkat Edilmesi Gerekenler

- ❑ Tablolarda gereksiz çizgi ve boşluk kullanmaktan kaçınılmalı
- ❑ Kalabalık tablo ve şekiller
- ❑ Renk ögesinin kullanımı
- ❑ Üretilen tablo/grafiğin hangi ortamda (basılı, web) sunulacağı
- ❑ Web ortamındaki sunumda etkileşim özelliği

Yararlanılan Kaynaklar

- ❑ Alaçam, Ö. (2010). Comparison strategies in different types of graphs. 24 Ekim 2014 tarihinde <http://etd.lib.metu.edu.tr/upload/12611699/index.pdf> adresinden erişildi.
- ❑ Carpendale, M.S.T. (2003). Considering visual variables as a basis for information visualisation. 24 Ekim 2014 tarihinde <http://innovis.cpsc.ucalgary.ca/innovis/uploads/Publications/Publications/Carpendale-VisualVariablesInformationVisualization.2003.pdf> adresinden erişildi.
- ❑ *Handbook of data visualization*. (2008). Chun-houh Chenn, Wolfgang Hardle ve Antony Unwin (eds.). Berlin: Springer.
- ❑ Huff, D. (1954). *How to lie with statistics*. New York: W.W. Norton & Company Inc.
- ❑ Tonta, Y. (1999). Bilimsel araştırmalarda istatistiksel tekniklerin kullanımı ve bulguların sunumu üzerine. *Türk Kütüphaneciliği*, 13(2), 112-124.
- ❑ Tonta, Y. (2011). Sosyal bilimlerde araştırma yöntemleri (açık ders malzemesi) Hacettepe Üniversitesi Edebiyat Fakültesi Bilgi ve Belge Yönetimi Bölümü. Ankara.
- ❑ UNESCO. (2012). Types of variables. 24 Ekim 2014 tarihinde http://www.unesco.org/webworld/idams/advguide/Chapt1_3.htm adresinden erişildi.
- ❑ Wills, G. (2012). Visualizing time: Designing graphical representations for statistical data. New York: Springer.