

# Projektarbete: Dataanalys

Gustav Hartvigsson  
Grebbebestads Folkhögskola (VT 2011)  
gustav.hartvigsson@gamil.com

13 februari 2011

# Innehåll

<b>I</b>	<b>Intro</b>	<b>3</b>
1	Projektets syfte	3
2	Frågeställning	3
3	Metod	3
4	Presentation	3
<b>II</b>	<b>main</b>	<b>4</b>
5	varför jag valde de tekniker som jag gjorde	4
5.1	Programmet . . . . .	4
5.2	Enkäten . . . . .	4
5.3	Bazaar . . . . .	4
6	Problem	5
6.1	Programmet . . . . .	5
6.1.1	Bibliotek . . . . .	5
6.1.2	Debugging . . . . .	5
6.2	Enkät . . . . .	6
<b>III</b>	<b>data och analys</b>	<b>6</b>
7	Rådata	6
8	Tabeller och grafer.	7
9	Analys av datan och jämförelse med annan statistik	9
10	Slutdiskussion	10
10.1	Erfarenheter . . . . .	10
10.2	Göra saker bättre . . . . .	10
10.3	Fortsatt forskning . . . . .	10
<b>IV</b>	<b>Förteckningar</b>	<b>11</b>

## Del I

# Intro

## 1 Projektets syfte

Syftet som jag hade i projektplanen var följande:

Att träna på dataanalys/lära sig hur man gör dataanalys.

En längre beskrivning kunde vara: *Att träna på dataanalys/lära sig hur dataanalys går till och se på olika metoder för att göra detta på.*

## 2 Frågeställning

Frågeställningen som jag hade i projektplanen är följande:

Se hur relationen mellan olika frågeställningar ser ut.

En längre och mer utförlig frågeställning skulle kunna vara: *Se hur relationen mellan olika svar är. Se till exempel hur många veganer som är för dödsstraff.*

## 3 Metod

Metoden beskrev jag i projektplanen är som följande:

Skriva en enkät med olika frågor som besvaras med kryssrutor.

Sedan stoppa in svaren i ett kalkylblad alt. en databas. där man kan göra förfrågningar (queries) och skapa sig en bild av de olika grupperna på skolan.

Jag är ganska nöjd med hur den är formuleringen. så jag anser inte att jag behöver ändra den.

## 4 Presentation

Jag hade en presentation 2011 02 07, denna presentation skrevs i  $\text{\LaTeX}$ [1] [2] och Beamer.

I Projektplanen skrev jag följande:

Göra en presentation som redogör för relationerna mellan olika svar. Även eventuellt göra ett program som kan göra queries på ett enkelt sätt.

Jag har tänkt att skriva presentationen i  $\text{\LaTeX}$  och visa upp programmet som jag skrev.

I den här rapporten tänkte jag framföra lite statistik som visar olika frågeställningar och deras relationer till varandra.

## Del II

# main

## 5 varför jag valde de tekniker som jag gjorde

### 5.1 Programmet

När jag tänkte på hur jag skulle lösa problemet kom jag att tänka på att det vore lätt att ha all data i en sqldatabas och sedan använda sql förfrågningar för att ta ut relationerna och räkna instanserna där det är sant.

Sedan behövde jag ett lätt använt språk att skriva en “front-end” i vilken jag skulle kunna använda den för att lägga in data och ta ut data med. Med dessa kriterier så blev valet Python 3.X [3] (även känt som python py3k eller Python3k) detta blev dock ett problem, mer om detta i sectionen “Problem” längre ner.

När det gäller valet av databas så var det inte så svårt, jag har arbetat med SQLite förut och kräver inte en bakändsserver vilket som jag skulle behöva använda om jag hade valt till exempel MySQL [4] eller PostgreSQL [5].

### 5.2 Enkäten

Enkäten skrev jag i klartext, dvs oformaterad och med bara standard tecken för att det skulle vara lätt att använda i Bazaar (se längre ner). och att det skulle vara lätt att redigera.

enkätens struktur är som följande:

Först kom en kort beskrivning av projectet

sedan en liten information om hur datan skulle användas

sedan frågorna:

Fråga 1:

Tror du på en/flera personlig(a) gud/gudar?

Ja [ ] Nej [ ]

ect ect.....

### 5.3 Bazaar

Jag använde Bazaar (även känt som bzt) som ett revisions kontroll system (RCV) för att hålla koll på koden och som ett slags av dagbok över ändringar jag har gjort.

Här kommer en beskrivning om vad Bazaar är för något från Bazaars hemsida:

Bazaar is a version control system that helps you track project history over time and to collaborate easily with others. Whether you're a single developer, a co-located team or a community of developers scattered across the world, Bazaar scales and adapts to meet your needs. Part of the GNU Project, Bazaar is free software sponsored by Canonical. [6]

All kod och databasen som har med det här projektet finns att hämta på min LaunchPad: <https://code.launchpad.net/~gustav-hartvigsson/+junk/Dataanalys>, intrunktioner finns på sidan om hur man hämtar koden och databasen.

## 6 Problem

När jag började på projektet hade jag en idé om att det skulle vara lätt att producera en enkät och sammanställa det i en databas, men verkligheten var inte så lätt. I den här sektionen kommer jag att berätta om de problem jag har haft under projektets gång.

### 6.1 Programmet

Jag hade lite problem när jag gjorde programmet som jag gjorde det, här kommer lite exempel på det som jag träffade på.

#### 6.1.1 Bibliotek

När jag valde Python visste jag att det fanns mängder av bibliotek som kunde användas för att rita grafer i, men jag tänkte inte på att de inte fanns i Python 3k. Detta resulterade i att jag inte kunde göra grafer i programmet som jag skrev. Jag gjorde en "utflykt" i konsten att porta från ett py2.x till py3k (se [7]), men jag lade ner det eftersom att jag inte förstod hur biblioteket fungerade. Det ända jag gjorde var en enkel portning så, trots att den var användbar. Så hade den inte alla funktioner som i originalet.

#### 6.1.2 Debugging

Jag utförde aldrig någon ordentlig debugging innan jag försökte lägga in svaren i databasen, detta resulterade i att programmet kraschade vid andra körningen av "add to db". Felet var att jag inte rensade strängen som jag sedan använde, för att lägga till data i databasen innan jag började på med att lägga till mer data till strängen.

Detta blev fixat i revition med namnet “ `gustav.hartvigsson@gmail.com-20110131122114-e6e1z0p12eadau65` ” (se `bzr glog` eller `bzr log`)<sup>1</sup>.

## 6.2 Enkät

När det kommer till enkäter så var det lite problem. Bland annat så var inte svarsrutorna för fråga 9 på samma sida som frågan.

Ett annat problem var att jag stavade fel på frågan 4 och 20, där jag skrev *sekelärt* istället för *sekulärt* och *morerna* istället för *moderna*.

Andra problem som var med själva frågeställningen var att vissa inte ville svara på med ett klart ja eller nej. Detta var ett problem som jag skulle ha förutsatt när jag gjorde frågorna, men det gjorde jag inte då.

Ett tredje problem som jag hade var att vissa inte svarade på vissa frågor. Mer om detta längre ner.

## Del III

# data och analys

Här kommer jag att skriva om hur data analysen gick till och hur jag löste vissa problem som uppstod, och visa någon fin graf som visar hur de olika relationerna ser ut.

## 7 Rådata

Jag fick in 44 svar på enkäten, vilka nu mera finns i en databas, som jag har lagt ut på min launchpad tillsammans med koden.

Så här ser ett stycke av databasen ut i csv format.

```
id,gudar,spoken,religos,sekulart,vegan,mars,globaluppvarmning [...]
1,0,0,0,1,0,1,1,1,0,0,0,1,0,1,1,1,1,0,1,1,0,0,1," "
2,0,0,0,1,0,0,1,1,1,1,1,1,1,0,1,1,1,0,1,1,0,1,1," "
3,0,0,0,0,0,0,1,1,0,1,1,1,1,1,1,1,1,0,1,1,0,1,1," "
4,0,0,0,1,0,1,1,1,1,1,1,0,1,0,1,1,1,0,1,1,0,1,0," "
5,0,0,0,1,0,1,1,1,0,0,0,1,1,0,1,1,1,0,0,0,1,1,0," "
6,1,1,0,1,0,1,1,1,0,1,1,1,1,0,1,1,1,0,1,1,0,1,0," Du skull [...]
```

Se `csv/awnsers2.csv` för datan i csv format eller `awnsers.sqlite` för datan i SQLite3 format <sup>2</sup>.

Datans struktur är inte så märkvärdig: 1 är santt och 0 är falskt.

---

<sup>1</sup>koden finns tillgänglig på min LaunchPad: <https://code.launchpad.net/~gustav-hartvigsson/+junk/Dataanalys>

<sup>2</sup>Se föregående fotnot

## 8 Tabeller och grafer.

I den här sektionen kommer jag att gå genom lite statistik även vilka kopplingar som finns mellan olika frågeställningar som var med i enkäten.

	Tror på gu(ar)	Tror på spöken	Är religiös	Är för ett sekulärt samhälle	Vegan eller vegetarian	Är för abort	Är för kvinnors rättigheter	Är för dödsstraff
Tror på gud(ar)		11	7	16	4	17	17	1
Tror på spöken	11		5	13	6	17	18	2
Är religiös	7	5		6	3	8	8	0
Är för ett sekulärt samhälle	16	13	6		6	35	35	2
Vegan eller vegetarian	4	6	3	6		8	9	0
Är för abort	17	17	8	35	8		42	4
Är för kvinnors rättigheter	17	18	8	35	9	42		4
Är för dödsstraff	1	2	0	2	0	4	4	
av totalt	18	19	9	35	9	42	43	4

Tabell 1: Ett litet urval av datan

Datan i tabellen är framställd via att använda `count` funktionen i programmet under `showDB`. Ett exempel på hur man kan göra: `(cmd) count vegan and dödsstraff` detta skriver ut siffran 0 på kommandoprompten.

Här kommer en tabell med procentsattser som är producerade genom  $\frac{cell}{tot_x}$  och  $\frac{cell}{tot_y}$ . Se tabell 2 på sidan 8.

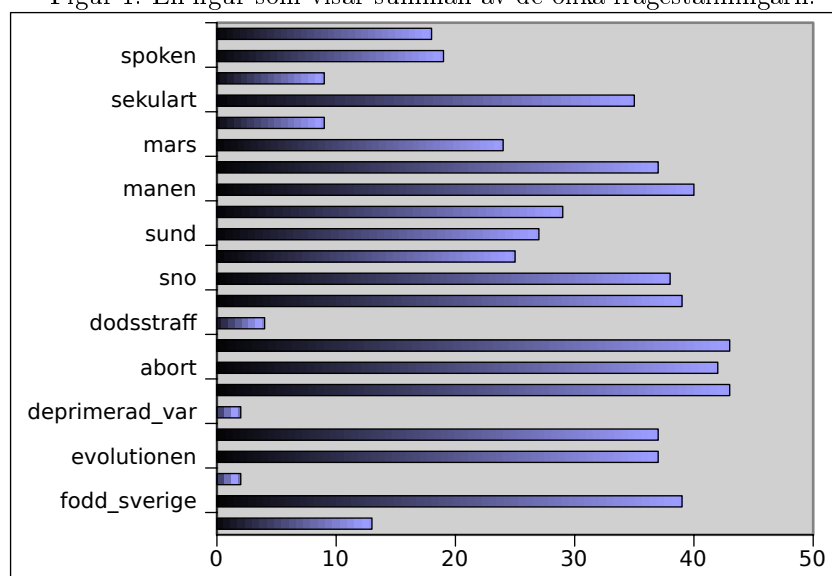
Utifrån datan så har jag sammanställt datan i ett diagram, som visar olika kolumner som har summerats, se figur 1 på sidan 8. alla etiketter är inte med i diagrammet, så här kommer etiketterna i samma ordning:

gudar, spoken, religiös, sekulärt, vegan, mars, globaluppvärmning, manen, motion, sund, deprimerad\_host, sno, oppet,dödsstraff, kvinnors\_rättigheter, abort, kvinnors\_kropp, deprimerad\_var, vetenskap, evolutionen, gast\_sverige, född\_sverige, föräldrar\_sverige

	Tror på gu(ar)	Tror på spöken	Är religiös	Är för ett sekulärt samhälle	Vegan eller vegetarian	Är för abort	Är för kvinnors rättigheter	Är för dödsstraff
Tror på gud(ar)	x	61%	38%	89%	22%	94%	94%	5,5%
Tror på spöken	61%	x	28%	68%	33%	89%	95%	11%
Är religiös	38%	28%	x	66%	33%	89%	89%	0%
Är för ett sekulärt samhälle	89%	68%	66%	x	17%	100%	100%	5%
Vegan eller vegetarian	22%	33%	33%	17%	x	89%	100%	0%
Är för abort	94%	89%	89%	100%	89%	x	100%	10%
Är för kvinnors rättigheter	94%	95%	89%	100%	0%	100%	x	9%
Är för dödsstraff	5,5%	11%	0%	5%	0%	10%	9%	x
av totalt	100%/18	100%/19	100%/9	100%/35	100%/9	100%/42	100%/43	100%/4

Tabell 2: samma data som i tabell 1 fast i procent.

Figur 1: En figur som visar summan av de olika frågeställningarn.





## 9 Analys av datan och jämförelse med annan statistik

I den här sektionen så kommer jag att framföra lite analys och jämförelse med andra källor som finns tillgängliga.

Vi kan börja att kolla på den statistik som vi har tillgänglig redan nu. Om vi tittar i tabell 1 på sidan 8 så kan vi se att den bara visar många som tror på vissa saker som även tror på något annat, och om vi tittar på figur 1 på sidan 8 detta är inte lätt att arbeta med. Men om vi tar samma urval och istället kollar på det i jämförelse med alla 44 svar som jag har fått in så får vi en bättre bild av vad gruppen som helhet tror och tycker, och göra en jämförelse med annan statistik, nämligen *Social values, Science and Technology* [8] från 2005.

%	Skolan	Special Eurobarometer 2005
tror på gudar:	18/44 $\approx$ 40%	23%
tror på spöken:	19/44 $\approx$ 43%	N/A
är för abort:	42/44 $\approx$ 95%	54% <sup>1</sup>

Tabell 3: En tabell som visar en jämförelse mellan skolan och sverige som helhet

Som vi kan se så är till exempel de som tror på gudar är något fler än vad som är riksgenomsnittet i Sverige (enligt *Social values, Science and Technology*).

<sup>1</sup>Frågan är dock felställd: *Protecting the dignity of any human unborn life*" vilket är svårtolkat och kan inte direkt översättas till den frågan jag ställde.

## 10 Slutdiskussion

Här så tänker jag visa de erfarenheter som jag har fått av att fullförja detta projekt, reflektera över hur jag skulle kunna göra saker bättre och ge min syn på hur man kan göra fortsatt forskning.

### 10.1 Erfarenheter

De erfarenheter jag har fått är som följer:

- Hur man gör enkäter, och hur man inte skall göra dem.
- Hur man sammanställer data på ett bra sätt.
- Att skriva rapporter i och brättas med  $\text{\LaTeX}$  egenheter.
- Att arbeta med mig själv och fullförja ett projekt.

### 10.2 Göra saker bättre

Om jag hade gjort projektet igen så skulle jag ha läst mer om hur man gör data analys och, hur man skall göra enkäter som är lätta att förstå och att man skall korrektur läsa hela enkäten innan man delar ut dem.

### 10.3 Fortsatt forskning

Om någon skall göra fortsatt forskning i detta ämne så är det viktigt att veta att det här är bara en liten demografisk grupp, och att det inte finns mycket data att gå på, i alla fall inte på SCB.

Om någon skulle göra ett liknande projekt som jag har gjort, så skulle man behöva ha en handledare som är kunnig i ämnet och som kan programera. Detta är dock inte något krav om man kan lite om ämnet i förhand och kan programera det simplaste programmen.

## Del IV

# Förteckningar

## Figurer

1	Summa av frågeställningar. . . . .	8
---	------------------------------------	---

## Tabeller

1	Ett litet urval av datan . . . . .	7
2	Procent: Ett urval av datan . . . . .	8
3	En tabell som visar en jämförelse mellan skolan och sverige som helhet . . . . .	9

## Referenser

- [1] various LaTeX project. *LaTeX – A document preparation system*. 1 febr. 2011. URL: <http://www.latex-project.org/>.
- [2] Wikimedia fundation various. “LaTeX - Wikipedia, the free encyclopedia”. I: (1 febr. 2011). URL: <http://en.wikipedia.org/wiki/LaTeX>.
- [3] various Python Foundation. *Python Programming Language – Official Website*. 31 jan. 2011. URL: <http://www.python.org>.
- [4] Wikimedia fundation various. “MySQL - Wikipedia, the free encyclopedia”. I: (31 jan. 2011). URL: <http://en.wikipedia.org/wiki/MySQL>.
- [5] Wikimedia fundation various. “PostgreSQL - Wikipedia, the free encyclopedia”. I: (31 jan. 2011). URL: <http://en.wikipedia.org/wiki/PostgreSQL>.
- [6] Various canonical. *Bazaar*. 1 febr. 2011. URL: <http://bazaar.canonical.com/en/>.
- [7] Gustav Hartvigsson och Gerald Kaszuba. *port of pygooglechart to py3k*. 5 jan. 2011. URL: <https://code.launchpad.net/~gustav-hartvigsson/+junk/pygooglechart-py3k>.
- [8] Special Eurobarometer European Commission. *Social values, Science and Technology*. Tekn. rapport. Juni 2005. URL: [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_225\\_report\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_225_report_en.pdf).