



Linköpings Tekniska Högskola
Campus Norrköping - ITN
TNM048 Informationsvisualisering
2005-03-06

i view

CAR DATA SET

VISUALISERING AV BILDATA

Alexander Petrovski
alepe261@student.liu.se

Frank van der Stelt
frava974@student.liu.se

Sammanfattning

Rapporten presenterar ett projekt i informationsvisualisering. Informationsvisualisering har som mål att presentera enormt stora mängder multivariat data på ett lättbegripligt sätt. Rapporten kommer att beskriva tillvägagångssättet från filen med den abstrakta datamängden till en applikation som kan hanterat dessa data. Förklaring av visualiseringskomponenter och interaktionsmetoder kommer att ske samt en motivering av val gällandes dessa. Gränssnittsdesignen kommer också att behandlas. En mängd bilder på applikationen visar resultatet på vårt arbete. Avslutningsvis kommer en diskussion med en beskrivning av framtida förbättringar av applikationen. Även en kortfattad användarmanual kommer även att presenteras.

Innehåll

1 INLEDNING	1
1.1 BAKGRUND	1
1.2 MÅL OCH SYFTE.....	1
1.3 METOD	1
<i>1.3.1 Utvecklingsmiljö och programmeringsspråk.....</i>	<i>1</i>
2 VISUALISERINGEN	2
2.1 DATA.....	2
2.2 VISUALISERING	2
<i>2.2.1 Parallella koordinater.....</i>	<i>2</i>
<i>2.2.2 Scatterplots</i>	<i>3</i>
<i>2.2.3 Statusbar.....</i>	<i>3</i>
2.3 INTERAKTION	4
<i>2.3.1 Filtrering av parallella koordinater.....</i>	<i>4</i>
<i>2.3.2 Filtrering av "scatterplots"</i>	<i>4</i>
2.4 GRÄNSSNITT	5
3 RESULTAT	7
3.1 FRAMTIDA FÖRBÄTTRINGAR.....	7
3.2 ANVÄNDARMANUAL.....	8
<i>3.2.1 Att komma igång/ avsluta.....</i>	<i>8</i>
<i>3.2.2 Att använda iVIEW</i>	<i>8</i>
KÄLLOR ANVÄNDA I PROJEKTET.....	9

1 Inledning

Rapporten beskriver ett projekt i informationsvisualisering, som är en del av civilingenjörsutbildningen i medieteknik.

1.1 Bakgrund

Många företag har idag enorma databaser fyllda med information om allt möjligt. Det är mycket svårt att ur dessa databaser få insikt om hur informationen hänger ihop. Chansen är stor att det gömmer sig mängder av viktiga samband mellan data. Data är oftast mycket abstrakt och för att lättare tillgodogöra sig informationen är det viktigt att presentera dessa data på ett överskådligt sätt. Informationsvisualisering är ett ämne som har just detta som mål. Med hjälp av avancerad datorgrafik och snabba datorer är det idag möjligt att effektivt kunna visualisera enormt stora datamängder. Det är möjligt att i realtid kunna interagera med data och informationsvisualisering blir allt viktigare i beslutsfattningen hos företag. Det är lätt att överblicka stora mängder data och med hjälp av statistiska medel kan data extrapoleras. På detta sätt kan beslutsfattningen underlättas och därigenom förbättras.

1.2 Mål och syfte

Ämnet informationsvisualiseringen ges som en del av visualiseringsprofilen på civilingenjörsutbildningen i medieteknik. Detta projekt är en del av kursen och målet är att fördjupa våra kunskaper i ämnet. Vi kommer att presentera vår applikation i denna rapport där det inledningsvis kommer att ske en beskrivning av data samt vilka visualiseringsmetoder som har används. Interaktionen med data beskrivs samt applikationens gränssnitt. Redovisning av resultat kommer att ske och i diskussionsdelen diskuteras förslag till framtida förbättringar av applikationen.

1.3 Metod

Ett antal böcker kring detta område finns att hitta och Internet har också en hel del information. Projektet har främst lutat sig mot boken *"Information Visualisation"* av Robert Spence samt föreläsninganteckningar för att på lämpligaste sätt visualisera datamängden. När det gäller programmeringen bör OpenViz dokumentationen nämnas som viktigaste källa av information. Flitigt användande av Internet för att hitta lösningar till problem som uppstod under programmeringsarbetet hjälpte till stor del till.

1.3.1 Utvecklingsmiljö och programmeringsspråk

Labborationerna i informationsvisualiseringen gjordes med hjälp av Microsoft Visual Studio .Net och OpenViz och det kändes naturligt att även göra projektet med hjälp av dessa produkter. Microsoft Visual Studio .Net är en utvecklingsmiljö som tillåter en att programmera i olika språk. Det är en mycket smidigt produkt som underlättar programmeringsarbetet rejält. OpenViz Particles är ett komponentbibliotek för informationsvisualisering. Det finns en mängd komponenter såsom olika grafer och olika tekniker för att manipulera samt interagera med data. Det är relativt lätt att skapa en användbar applikation som är lätt att hantera och som dessutom ser bra ut.

2 Visualiseringen

2.1 Data

Datamängden som visualiserades består av 10 datadimensioner med bildata. Med en datadimension menas en kolumn med data i excelformat. 154 bilar med 10 datadimensioner skulle på lämpligt sätt visualiseras. Det första som gjordes var en granskning av datamängden. Vissa datadimensioner ansågs vara viktigare än andra. Datadimensioner såsom bilmärke och modell har därför inte visualiserats och datamängden kan därmed inte filtreras med avseende på dessa. För att visa ursprungsland, färglades data i olika färger. De övriga datadimensioner är numeriska och lämpar sig bättre för visualisering; dessa är mera intressanta att titta på för de flesta.

mpg	cylinders	horsepower	accel	year	weight	origin	make	model	price
15.5	4	43	21.5	78	1985	GERMANY	Volkswagen	1200 Hatch	2400
16.2	4	43	14.4	78	1800	USA	Ford	181	1900
16.5	4	43	19.4	78	1985	JAPAN	Mazda	200-SX	2200
16.9	4	52	18.6	78	2070	JAPAN	Datsun	200SX	2725
17	4	52	16.4	78	1800	JAPAN	Honda	210	2250
17	8	89	15.5	78	3365	USA	Oldsmobile	210	3300
17.5	8	80	13.2	78	3735	USA	Dodge	210	3125
17.6	8	80	12.8	78	3570	USA	Mercury	2400	2850
17.6	6	60	19.2	78	3535	USA	Pontiac	264L	2800
17.7	6	62	18.2	78	3155	USA	Chevrolet	280-2K	3275
18.1	6	62	15.8	78	2965	USA	Ford	3000	2375
18.1	4	63	15.4	78	2720	USA	Ford	310 GX	2275
18.2	6	63	17.2	78	3430	USA	Plymouth	310	2700
18.5	6	64	17.2	78	3210	USA	AMC	4000	2300
18.6	6	65	15.8	78	3380	USA	Buick	5000	3300
19.1	6	65	18.7	78	3070	USA	Mercury	5000S DL	2425
19.2	6	65	18.7	78	3520	USA	Dodge	504	2700
19.2	6	65	15.1	78	3410	USA	AMC	5085 DL	2425
19.2	8	65	13.2	78	3425	USA	Chevrolet	510 Hatch	3900
19.4	6	65	13.4	78	3445	USA	Buick	510	4400
19.4	8	65	11.2	78	3205	USA	Ford	604SL	2525
19.8	8	65	13.7	78	4080	USA	Dodge	626	3000
19.9	4	65	16.5	78	2155	USA	Chevrolet	626	2100
20.2	4	66	14.2	78	2560	JAPAN	Toyota	810 Maxima	2975
20.2	4	67	14.7	78	2300	JAPAN	Datsun	900S	2775
20.2	4	67	14.5	78	2230	USA	Dodge	990LE	2250
20.2	4	67	14.8	78	2515	JAPAN	Toyota	Accord LX	3700
20.3	4	67	16.7	78	2745	USA	Plymouth	Accord	3200
20.5	4	67	17.5	78	2855	USA	Oldsmobile	Accord	2400
20.5	4	67	14.9	78	2405	JAPAN	Datsun	Aries SE	2975
20.6	5	67	15.9	78	2830	GERMANY	Audi	Aries SM	4475
20.6	6	67	13.6	78	3140	SWEDEN	Volvo	Aspen 6	5875
20.8	4	68	15.7	78	2795	SWEDEN	Saab	Aspen	4200
21.1	6	68	15.8	78	3410	FRANCE	Peugot	Aspen	5450
21.5	4	68	14.9	78	1990	GERMANY	Volkswagen	B210 GX	3675
21.6	4	68	16.6	78	2135	JAPAN	Honda	Camaro	3425
22	6	68	15.4	78	3245	USA	Pontiac	Cavariae CL	3325
22.3	6	68	18.2	79	2990	USA	Mercury	Cavalier	3200
22.4	4	70	17.3	79	2890	USA	Ford	Cavalier2D	2975
23	6	70	18.2	79	3265	USA	AMC	CavalierSM	3150
23	6	70	18.6	79	3360	USA	Dodge	Celica GT	3325

Fig.1

En del av datamängden

2.2 Visualisering

OpenViz har en mängd olika metoder för visualisering av data. Vissa metoder är mera sofistikerade än andra men utgångspunkten är att det ska vara lätt att tyda och manipulera datamängden.

2.2.1 Parallella koordinater

Parallella koordinater¹ är en enkel, kraftfull och effektiv geometrisk visualiseringsteknik i 2D för att presentera, analysera, filtrera och tolka flerdimensionell data. Varje vertikal linje representerar en dimension i datamängden och representeras med en axel. Varje axel visar datamängdens respektive maximum - och minimumvärden och varje linje motsvarar en bil ur datamängden. Punkten där linjen skär axeln, motsvarar bilens aktuella värde en specifik dimension. Alla axlar är separat skalade så att alla värden i dimensionerna kommer med.

¹ Parallella koordinater presenteras av Alfred Inselberg 1981 som ett nytt sätt att presentera flerdimensionell information.

Parallella koordinater

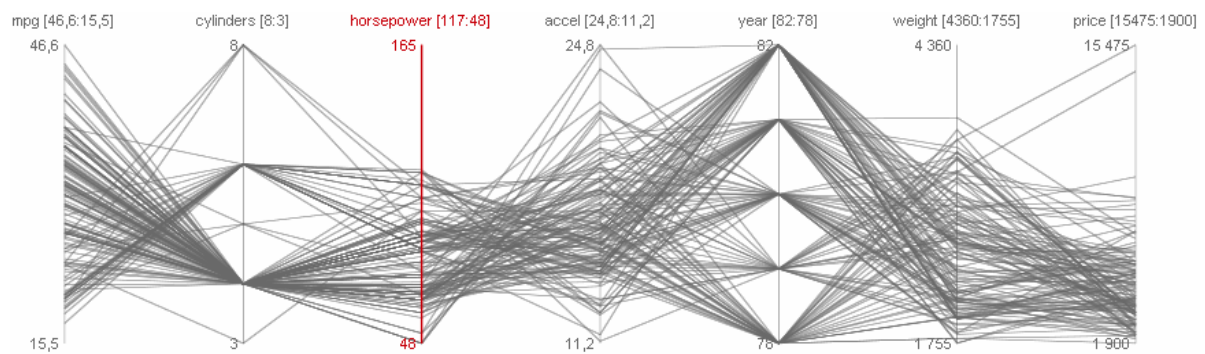


Fig.2

Bilden visar datamängden och varje vertikal linje (axel) representerar en dimension. Varje dimension har sitt namn utskrivet tillsammans med urvalets maximum – och minimumvärden.

2.2.2 Scatterplots

För att enkelt upptäcka samband mellan data användes två stycken tvådimensionella scatterplots. Två dimensioner valdes eftersom det är lättare att få en uppfattning om det verkligen finns en korrelation. Det är svårare att upptäcka samband i tre dimensioner. Förutom att "parallel coordinates" används för att åstadkomma "overview and detail" kan det väljas vilka datadimensioner x- och y-axlarna ska representera. På så sätt kan gå det att välja samma datadimensioner på båda scatterplots. Detta gör det möjligt att manipulera den högra grafen under tiden som översikten på den vänstra grafen bibehålls.

En "rangerscrollbar" lades till på vänstra grafen och är länkad till den högra grafen.

Den högra grafen har en så kallad "dragrectangle interactor" som möjliggör en förstoring av ett område som användaren specificerar. "Scatterplots" är ett naturligt sätt att visualisera tvådimensionell data. Det behövs inga specialister för att tyda datamängden utan det är ett intuitivt sätt att som presentation.

Ursprungslandet valde vi att färglägga. En legend visar vilken färg som tillhör vilket land men filtrering med avseende på ursprungslandet är inte möjligt.

2.2.3 Statusbar

En "statusbar" längst ner på skärmen visar upp instruktioner och extra data. En obligatorisk del av projektuppgiften var att ha med statistik i applikationen. Det är svårt att veta vilka statistiska värden i datamängden som är väsentliga. Därför valdes det att beräkna genomsnittsvärden per datadimension.

När en axel markeras i parallella koordinater, beräknas genomsnittet samt att min- och maxvärdena visas. Det visas även hur många bilar som ingår i ett intervall efter filtrering och vilken axel som använts för att utföra filtreringen. Vilken datafil som används i visualiseringen visas också.



Fig.3

Statusbaren

2.3 Interaktion

Det är viktigt att interaktionen sker på ett intuitivt sätt. Valet av visualiseringskomponenter begränsar antalet interaktionsmetoder som kan användas. Det finns bara ett sätt att filtrera datamängden när parallella koordinater används. Valet är lite större när det gäller "scatterplots" men fortfarande begränsat. Applikationen har en sorts hierarkisk, envägsänkning. Parallella koordinater är länkade till båda "scatterplots". Den vänstra grafen är länkad till den högra grafen och den högra grafen är inte länkad till någonting alls; detta för att åstadkomma "overview and detail". Man går ifrån en generell presentation av all data till en mer specifik presentation som beror på användarens intresse eller nyfikenhet.

2.3.1 Filtrering av parallella koordinater

För att filtrera datamängden krävs det att interaktion av användaren. Att interagera med de parallella koordinaterna utförs genom att hålla inne vänster musknapp i närheten av valfri axel samt dra uppåt eller nedåt. Då filtreras alla bilar som inte möter dem krav på maxvärden som valts. För att filtrera datamängdens minimumvärden istället, skall användaren trycka på [SHIFT] tangenten och därefter dra uppåt eller ner med hjälp av musen samtidigt som vänster musknapp hålls inne.

Filtrering av datamängden

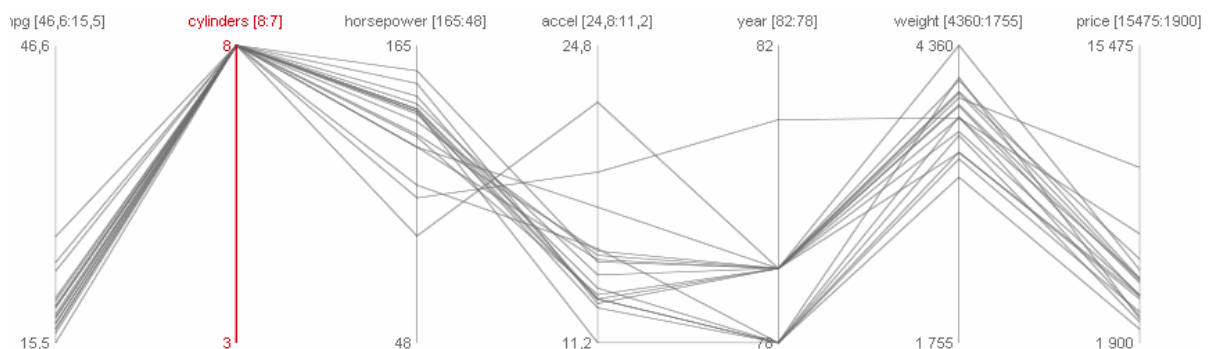


Fig.4

Interaktion med parallella koordinater kan vara ett kraftfullt verktyg för beslutsfattningen i företag. Att kunna hantera många dimensioner simultant kan vara till god hjälp. Här har användaren valt att inkludera endast bilar med åtta cylindrar.

2.3.2 Filtrering av "scatterplots"

Ett antal interaktionssätt har redan nämnts. "Comboboxar" valdes för att ändra x- och y-koordinater på båda graferna. Inte alla datadimensioner är intressant. Bara numeriska datadimensioner kan väljas. För att kunna filtrera datamängden har en "rangescrollbar" används som filtrerar datamängden med avseende på x-axeln i vänstra grafen. Detta kan vara bra om ett visst intervall eller t ex ett visst år. "Rangescrollbaren" är bara länkad till den högra grafen och inte till "parallel coordinates" grafen.

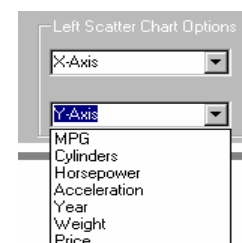


Fig.5 - En combobox

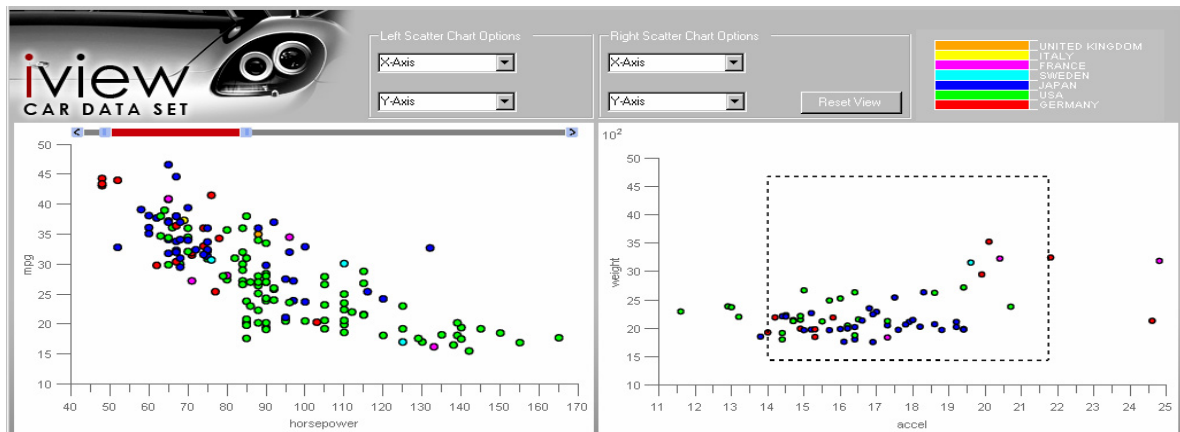


Fig.6

TVå scatterplots

Tanken var att göra huvudfiltreringen med hjälp av parallella koordinater och att sedan göra ytterliggare en filtrering i den vänstra grafen. Resultatet presenteras i den högra grafen. En "dragrectangle interactor" har länkats till den högra grafen och därigenom går det att göra förstoringar av ett visst område. Man kan komma tillbaka till originalvyn genom knappen "reset view".

När ett dataelement klickas i graf två, kommer all information om elementet att synas längst ner på högra sidan i "Car Information Box".

Car Information Box	
mpg =	29,9
cylinders =	4
horsepower =	65
acceleration =	20,7
year =	81
weight =	2380
origin =	USA
make =	Ford
model =	Escort 2H
price =	4100

Fig.7 - Bilinformationen

2.4 Gränssnitt

IVIEW har ett lättanvänt gränssnitt som en användare med ordinär datorvana känner igen.

Programmet har rullgardinsmenyer med två huvudalternativ, "file" och "help". Under "file" går det att öppna en fil ("open") i formatet Excel 8.0 samt avsluta programmet med hjälp av "quit". Under "help" finns det ett menyval "about" som ger information om projektet (programmet).

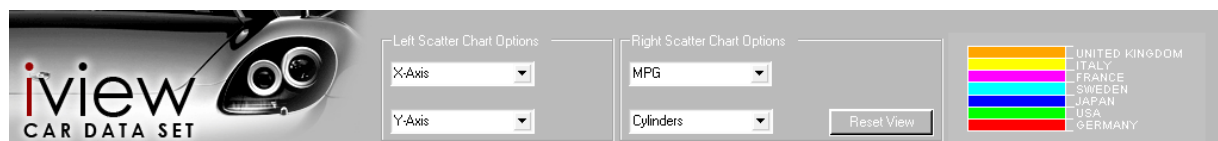
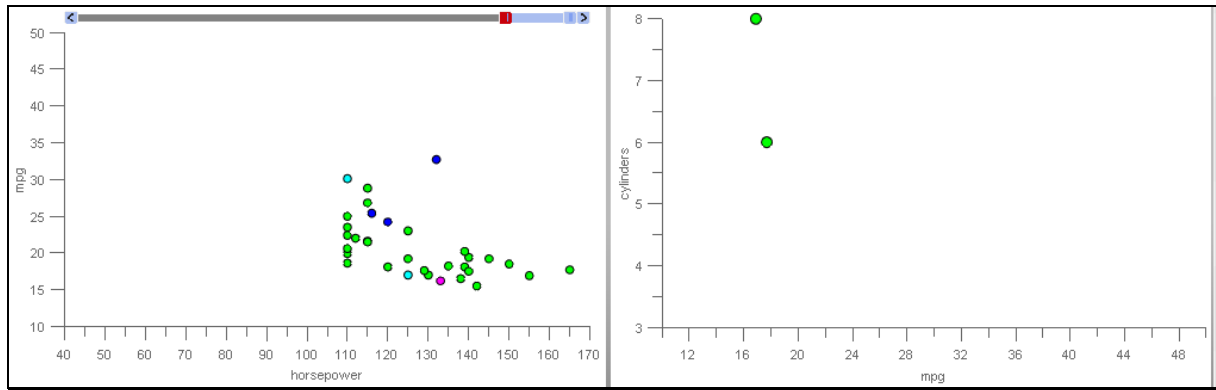


Fig.8

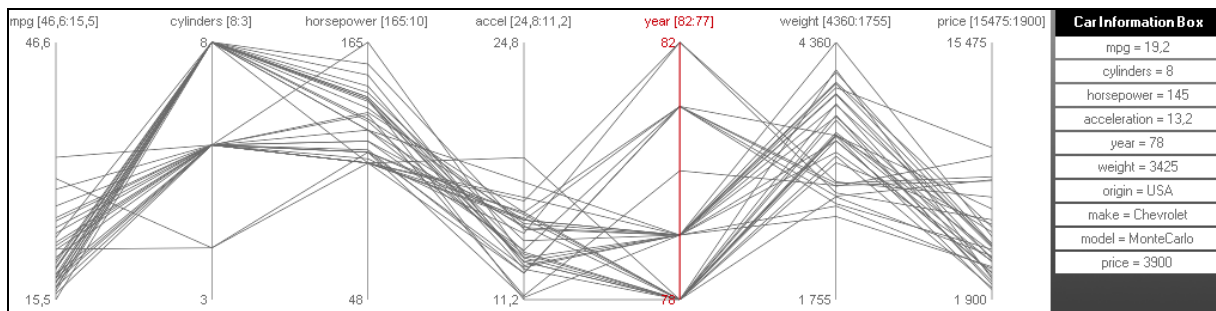
Programmets övre del

Längst upp till vänster finns programmets logotyp och till höger finns det fyra stycken "comboboxar" med titlarna "Left Scatter Chart Options" och "Right Scatter Chart Options". Dessutom finns det en knapp med texten "Reset View".

**Fig.9**

Programmets mellandel

Längst ut till höger finns det en legend som visar vilka färger som hör till vilket land i datamängden. I mitten finns de två scatterplottarna intill varandra. Rangescrollbaren i den vänstra filtrerar datat och visar upp det i det högra scatterplotten.

**Fig.10**

Programmets undre del

Längst ner i programfönstret finns representationen av de parallella koordinaterna. I ”statusbaren” allra längst ner finns det information om antal bilar, vald axel, min/max-värden, medelvärde samt sökväg till filen som öppnats.

3 Resultat

Bilder av resultatet har redan presenterats i den löpande texten. Det har blivit en stabil och flexibel applikation som på ett enkelt sätt presenterar bildata. Det är lätt att filtrera datamängden utan att tappa bort sig i för många filtreringsalternativ. Datamängden som visualiseras är förhållandevis liten och det skulle vara intressant att se hur applikationen presterar när en datamängd används som är 1000 gånger större. Det är mycket vanligt att datamängderna storleksmässigt uppnår giga- eller terabyte.

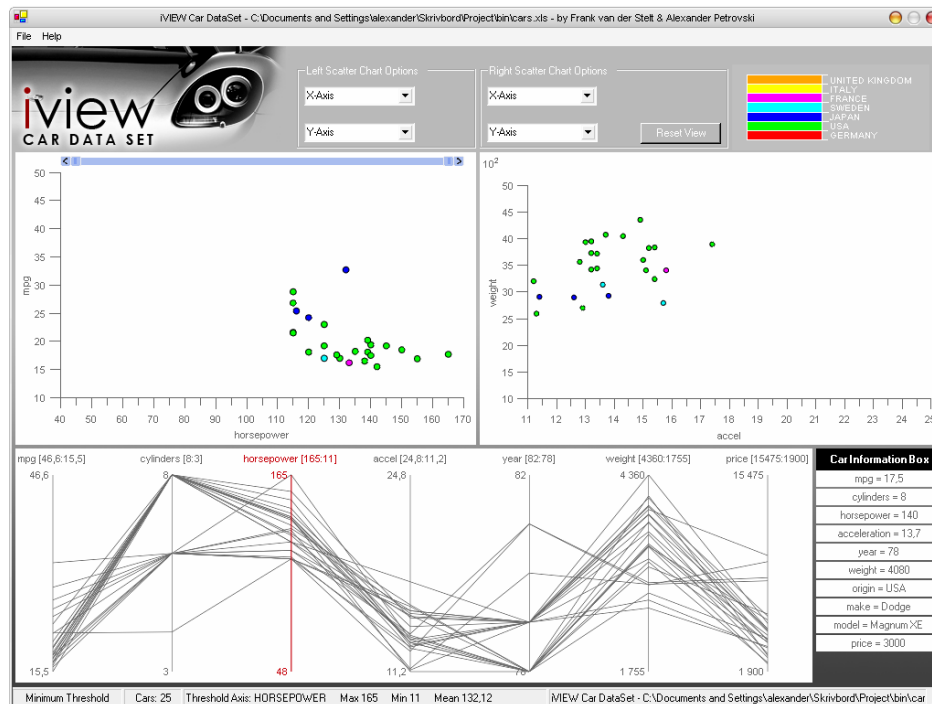


Fig.11
Resultatet

3.1 Framtida förbättringar

Såsom i många andra projekt inleds det med stora idéer, förhoppningar och programmeringsarbetet sätter igång. Efter en vecka brukar idéer oftast begränsas eftersom problem har stötts på som inte verkar lösbara och som tar upp all tid som finns. Arbetet i detta projekt började på detta sätt men efter en vecka började saker och ting fungera som det var tänkt. Innan programmeringsarbetet startades, gjordes en kort litteraturstudie och gjorde en skiss av applikationen.

Den slutliga applikationen har allt vi hade tänkt oss och lite till. Dessvärre går förbättringar alltid att göra. Vi skulle gärna markera ursprungsländerna i våra parallella koordinater med hjälp av samma färgskala som vi har använd oss av i de övriga grafer. Ett användarval av dimensioner i våra parallella koordinater skulle kunna vara en förbättring - det skulle i alla fall öka applikationens flexibilitet.

En resetknapp för vår "rangescrollbar" är också något skulle kunna läggas till.

Filtrering på ursprungsland är i nuvarande applikation omöjligt och detta skulle säkerligen läggas till om tiden funnits. Detta skulle kunna utföras med hjälp av en världskarta.

Istället för att ha en "rangescrollbar" på vänstra grafen och en "dragrectanglemanipulator" på den högra kunde man ha valt att enbart ha en vanlig "rectanglemanipulator" på den vänstra grafen. Med hjälp av en sådan kunde man förstora ett område och vissa upp det på den högra grafen - en elegantare lösning.

Förutom tidigare nämnda förbättringar går det alltid tänka på extra flexibilitet genom att användaren själv kan bestämma ländernas färg eller ändra textens typsnitt. Dessa trivialiteter ansågs dock inte ha så mycket med visualiseringsprojektet att göra utan lämnades därhän.

3.2 Användarmanual

Systemkrav är att en operativsystemet Windows 2000 eller Windows XP tillsammans med en OpenViz Particles viewerlicens. Datorn bör vara utrustad med minst 256mb minne samt ha en processor på minst 2Ghz (eller motsvarande). Minimikrav på skärmupplösning är 1024x768 pixlar.

3.2.1 Att komma igång/avsluta

För att starta programmet skall applikationen iVIEW.exe exekveras. För att avsluta programmet finns det två sätt att tillgå; antingen klicka på kryssset längst upp till höger eller gå till [File] i rullgardinsmenyerna och klicka på [QUIT]

3.2.2 Att använda iVIEW

När programmet laddats färdigt, är följande användarscenario ett exempel på hur programmet skulle kunna utnyttjas.

1. Gå till [FILE > OPEN]
2. Dubbelklicka på *cars.xls*
3. Gå till dem parallella koordinaterna och klicka på [SHIFT]
4. För musen till axeln med namnet *horsepower*
5. Var nog med att muspekaren står på nedre halvan av axeln, håll inne vänster musknapp och dra uppåt längs axeln
6. Notera att datat filtreras och släpp vänster mus i önskad position
7. Gå till vänstra grafen och använd "rangescrollbaren" för att ytterligare göra en filtrering
8. Filtreringen visas i högra grafen
9. Gå till comboboxen där det står "Right Scatter Chart Options" och klicka på menyvalet "X-Axis". Välj där "MPG" samt "Cylinders" på "Y-Axis"
10. Gå till vänstra grafen och prova att interagera med "rangescrollbaren" igen.
11. Gå till högra grafen och klicka på någon av de runda cirklarna (bilarna). Nu visas bilens data i "Car Information Box"

Källor använda i projektet

Böcker:

Robert Spence, *Information Visualization*, 2001, Addison-Wesley

Stefan Arvidsson och Jesper Ek, *Programmera med Visual Basic 6*, 2000, Pagina Förlag

Hemsidor:

<http://forum.av.s.com/2/OpenTopic> (2005-03-09)

<http://msdn.microsoft.com/> (2005-03-09)

<http://servus.itn.liu.se/courses/tnm048/> (2005-03-09)

http://servus.itn.liu.se/smardoc/project_results/knowledge/HTE/summary.html (2005-03-09)