



<第3回>
データ可視化とツール

統計数理研究所

中野 純司

nakanoj@ism.ac.jp

データ可視化とツール

概要

データサイエンティストとしてデータ分析を行う際に必要な可視化の考え方と、それを実行するためのフリーソフトウェアを紹介する

1. はじめに
2. 静的なグラフィックス
3. 動的なグラフィックス
4. 対話的なグラフィックス



1. はじめに

データ可視化とは

- データ(大量で複雑、その生成過程に興味あり)の特徴を、**目視で直観的に**捉えられるようにすること
- 簡単なものは「グラフ」
 - 円グラフ、棒グラフ、折れ線グラフ(初等教育)
 - ヒストグラム、散布図(中等教育)
- ここでは計算機によるより複雑な「グラフィックス」について説明する
 - 「グラフ」の延長、**散布図行列、3次元散布図、射影追跡、平行座標プロット、モザイクプロット**

データ可視化の目的

- データそのものや解析結果をわかりやすく表示する
 - ～プレゼンテーション
 - ～Excelのグラフ機能
- 対話的にグラフィックスを操作しながらデータの特徴を探る
 - ～探索的データ解析
 - ～ビジュアルデータマイニング

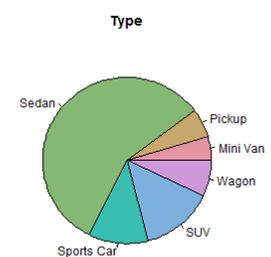
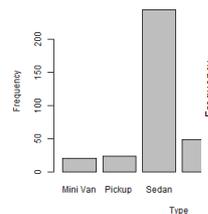
2. 静的なグラフィックス

(代表的な) 静的なグラフィックス

- **カテゴリ変数(色や地区など)データの集計結果**

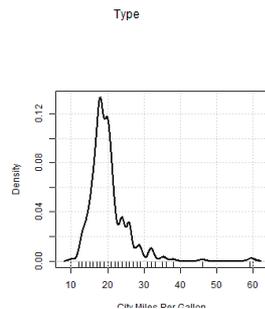
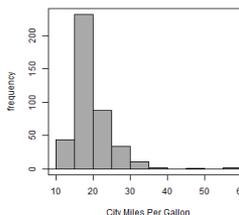
- **棒グラフ、円グラフ**

- ただし、並べることで複数の個体データを比較するためにも使える



- **連続変数(温度や身長など)データの分布**

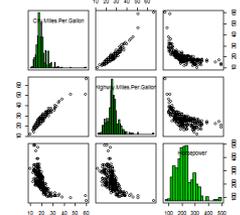
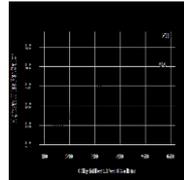
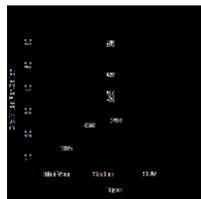
- **ヒストグラム、それをスムーズにしたもの(カーネル密度推定など)**



- **箱ひげ図(主として複数の集団を比較するに用いられる)**

- **二つの連続変数データの分布**

- **散布図**



- これを複数の変数のペアにしたもの: **散布図行列**

静的なグラフィックス描画における注意

- 必要な情報のみを示し、誤解を与えないこと
 - 円グラフや棒グラフを「立体化」することは誤解を与え易いので勧められない
 - 色の使い方は重要
 - 明度の高い膨張色（赤、橙、黄色のような暖色系の色）と収縮色（青、青緑のような寒色系の色）は大きさに対して誤解を与える
 - 区別するためには「パステルカラー」がよい

静的なグラフィックスのためのソフトウェア

- 静的なグラフィックスを描くため(および統計解析のため)にはフリーソフトウェアRがよい
 - <http://www.r-project.org/>
 - “R is a free software environment for statistical computing and graphics”
- Rは最もよく利用されているフリーの統計解析ソフトウェアで、グラフィックス機能が豊富、日本語の参考書も多い(例えば、参考文献 2)
- 簡単にそれなりの図が描けるし(高機能命令)、思い通りのグラフィックスを描くこともできる(低機能命令を用いてプログラミングが必要)
- パッケージ R Commander を使うとさらに簡単に使える



Windows taskbar and sidebar area containing various application icons and system information.

- Taskbar icons: R Commander, Firefox, Internet Explorer, Google Chrome, LibreOffice, TeXworks, and system tray icons (volume, network, power).
- Sidebar text: 編集, 自由なソフト, の条件に従え, 条件の詳細に, 多くの貢献者, は 'contr', また R や R の, citation(), 'demo()' と入, help.start(), 'Se...' と入力す, R..., R..., 要求されたパッケ, Rcmdrのバージョン, > |, < ▶

R コマンダー

ファイル 編集 データ 統計量 グラフ モデル 分布 ツール ヘルプ

データセット: データセットの編集 データセットを表示 モデル:

Rスクリプト Rマークダウン

Empty script editor area.

出力 実行

Empty output console area.

メッセージ

```
[1] メモ: Rコマンダーのバージョン 2.0-3: Wed Jan 29 14:09:27 2014
```


ファイルまたはクリップボード、URL か...

データセットの編集

データセットを表示

モデル: Σ <アクティブモデルなし>

データセット名を入力: Dataset

ファイル内に変数名あり:

欠測値の記号: NA

データファイルの場所

 ローカルファイルシステム クリップボード インターネットの URL

フィールドの区切り記号

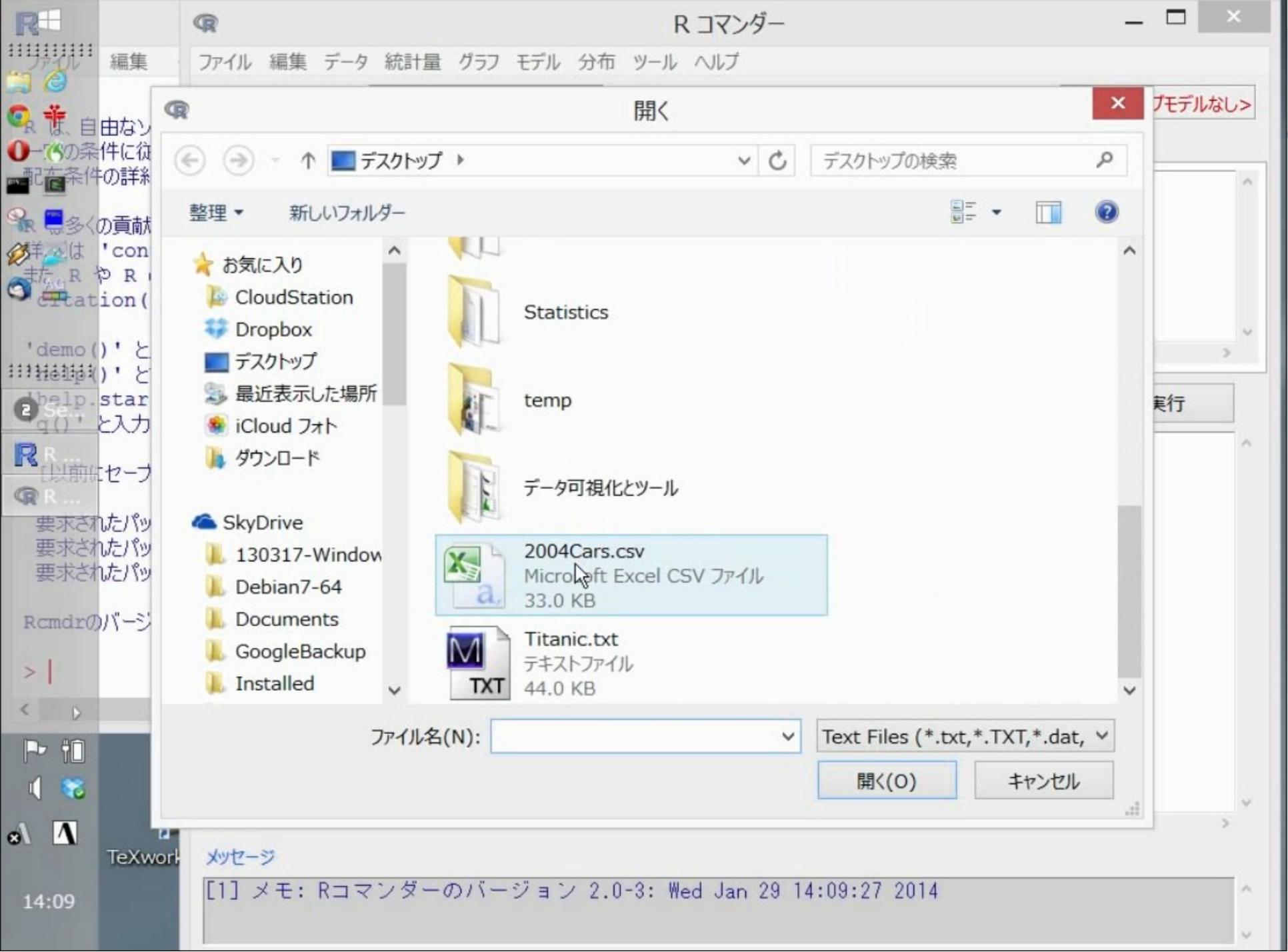
 空白 カンマ タブ その他 指定:

小数点の記号

 ピリオド[.] カンマ[,] ヘルプ OK キャンセル 実行

メッセージ

[1] メモ: Rコマンダーのバージョン 2.0-3: Wed Jan 29 14:09:27 2014



開く

デスクトップ ▶ デスクトップの検索

整理 ▼ 新しいフォルダー

お気に入り

CloudStation

Dropbox

デスクトップ

最近表示した場所

iCloud フォト

ダウンロード

SkyDrive

130317-Window

Debian7-64

Documents

GoogleBackup

Installed



Statistics



temp



データ可視化とツール



2004Cars.csv
Microsoft Excel CSV ファイル
33.0 KB



Titanic.txt
テキストファイル
44.0 KB

ファイル名(N):

Text Files (*.txt,*.TXT,*.dat, ▼

開く(O)

キャンセル

R コマンド

ファイル 編集 データ 統計量 グラフ モデル 分布 ツール ヘルプ

データセット: Dataset データセットの編集 データセットを表示 モデル: <アクティブモデルなし>

Rスクリプト Rマークダウン

```
Dataset <- read.table("C:/Users/nakanoj/Desktop/2004Cars.csv",  
  header=TRUE, sep=";", na.strings="NA", dec=".", strip.white=TRUE)
```

出力

実行

```
> Dataset <- read.table("C:/Users/nakanoj/Desktop/2004Cars.csv",  
+ header=TRUE, sep=";", na.strings="NA", dec=".", strip.white=TRUE)
```

メッセージ

```
[1] メモ: Rコマンドのバージョン 2.0-3: Wed Jan 29 14:09:27 2014  
[2] メモ: データセット Dataset には 428 行、14 列あります。
```

	Vehicle.Name	Type	Drive
409	Chevrolet Silverado SS	Pickup	AWD
410	Chevrolet SSR	Pickup	rear
411	Dodge Dakota Regular Cab	Pickup	rear
412	Dodge Dakota Club Cab	Pickup	rear
413	Dodge Ram 1500 Regular Cab ST	Pickup	rear
414	Ford F-150 Regular Cab XL	Pickup	rear
415	Ford F-150 Supercab Lariat	Pickup	AWD
416	Ford Ranger 2.3 XL Regular Cab	Pickup	rear
417	GMC Canyon Z85 SL Regular Cab	Pickup	rear
418	GMC Sierra Extended Cab 1500	Pickup	rear
419	GMC Sierra HD 2500	Pickup	AWD
420	GMC Sonoma Crew Cab	Pickup	AWD
421	Mazda B2300 SX Regular Cab	Pickup	rear
422	Mazda B4000 SE Cab Plus	Pickup	AWD
423	Nissan Frontier King Cab XE V6	Pickup	AWD
424	Nissan Titan King Cab XE	Pickup	AWD
425	Subaru Baja	Pickup	AWD
426	Toyota Tacoma	Pickup	rear
427	Toyota Tundra Regular Cab V6	Pickup	rear
428	Toyota Tundra Access Cab V6 SR5	Pickup	AWD



データセット: Data データセットを表示 モデル: <アクティブモデルなし>

Rスクリプト Rマークダウン

```
Dataset <- read.table(
  header=TRUE, sep=",",
  fix(Dataset)
```

```
"/Desktop/2004Cars.csv",
dec=".", strip.white=TRUE)
```

実行

出力

```
> Dataset <- read.tabl
+ header=TRUE, sep=
> fix(Dataset)
```

```
"/Desktop/2004Cars.csv",
, dec=".", strip.white=TRUE)
```

- 色パレット...
- インデックスプロット...
- ヒストグラム...
- 密度推定...
- 幹葉表示...
- 箱ひげ図...
- QQプロット...
- 散布図...
- 散布図行列...
- 折れ線グラフ...
- 条件付き散布図
- 平均のプロット...
- ドットチャート...
- 棒グラフ...
- 円グラフ...
- 3次元グラフ ▶
- グラフをファイルに保存 ▶

メッセージ

[2] メモ: データセット Dataset には 428 行、14 列あります。
[3] メモ: データセット Dataset には 428 行、14 列あります。

Windows taskbar and desktop icons including Start menu, taskbar, and various application icons like Firefox, LibreOffice, and TeXwork.

R コマンドー

ファイル 編集 データ 統計量 グラフ モデル 分布 ツール ヘルプ

棒グラフ

変数 (1つ選択)
Drive
Type
Vehicle.Name

ラベルを表示
x 軸のラベル <auto>
y 軸のラベル <auto>
グラフのタイトル <auto>

ヘルプ リセット **OK** キャンセル 適用

実行

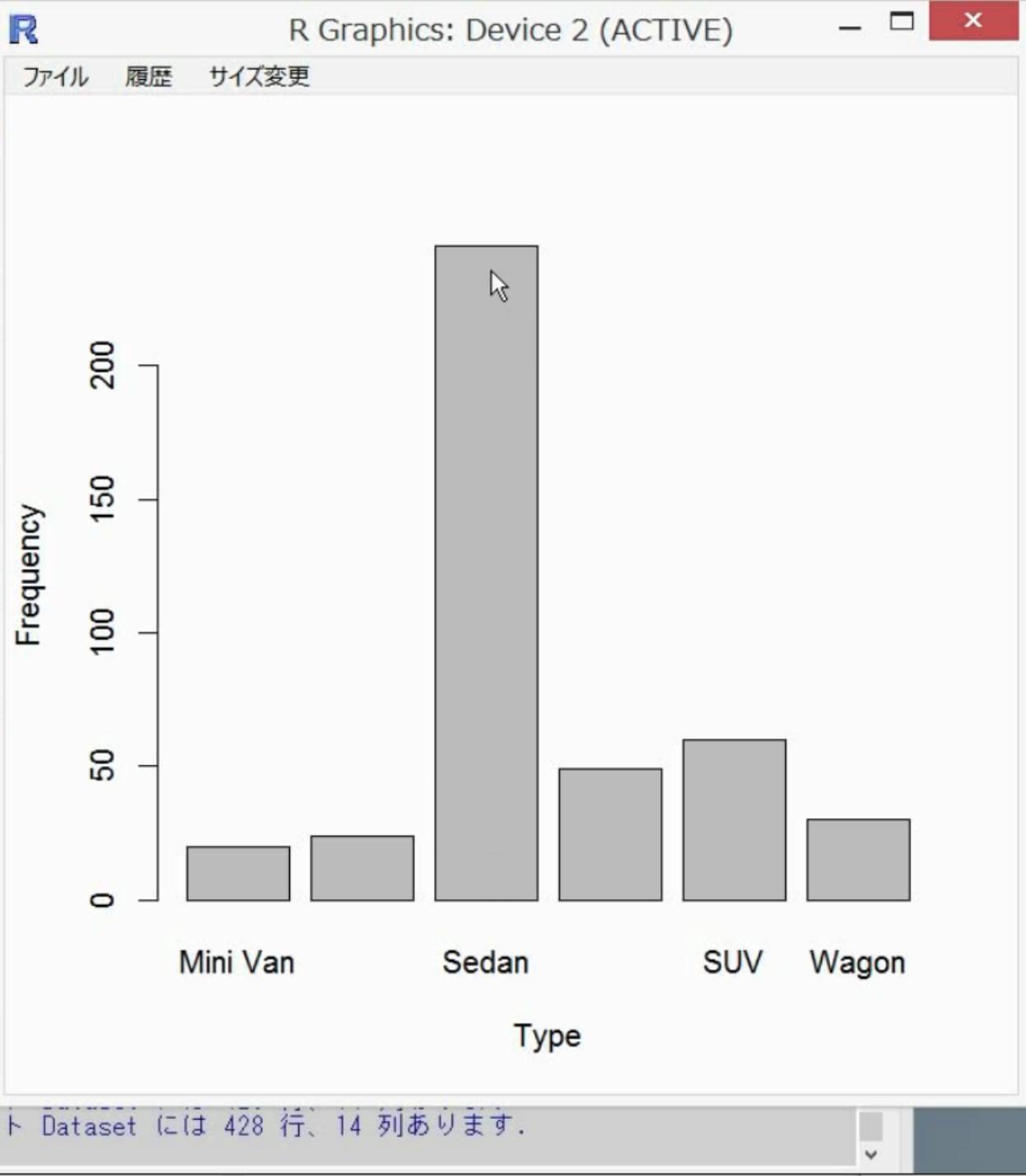
```
> Dataset <- read.table("C:/Users/nakanoj/Desktop/2004Cars.csv",
+ header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
> fix(Dataset)
```

メッセージ

```
[2] メモ: データセット Dataset には 428 行、14 列あります。
[3] メモ: データセット Dataset には 428 行、14 列あります。
```

Windows taskbar and sidebar icons including Start, Internet Explorer, Firefox, and various application icons.

R GUI menu and command window. Menu items: ファイル, 編集, データ, 統計. Command window content:
データセット: Dataset
Rスクリプト Rマークダウン
Dataset <- read.table(...)
header=TRUE, sep='...'
fix(Dataset)
barplot(table(Dataset))
出力
> Dataset <- read.table(...)
> header=TRUE, sep='...'
> fix(Dataset)
> barplot(table(Dataset))
メッセージ
[2] メモ: データセット Dataset には 428 行、14 列あります。



ファイル 編集 データ 統計量 **グラフ** モデル 分布 ツール ヘルプ

データセット: Data

データセットを表示

モデル: <アクティブモデルなし>

Rスクリプト Rマークダウン

```
Dataset <- read.table(
  header=TRUE, sep=","
fix(Dataset)
barplot(table(Dataset$
```

```
esktop/2004Cars.csv",
dec=".", strip.white=TRUE)
ylab="Frequency")
```

実行

出力

```
> Dataset <- read.tabl
+ header=TRUE, sep=
> fix(Dataset)
> barplot(table(Datase
```

```
/Desktop/2004Cars.csv",
, dec=".", strip.white=TRUE)
", ylab="Frequency")
```

色パレット...

インデックスプロット...

ヒストグラム...

密度推定...

幹葉表示...

箱ひげ図...

QQプロット...

散布図...

散布図行列...

折れ線グラフ...

条件付き散布図

平均のプロット...

ドットチャート...

棒グラフ...

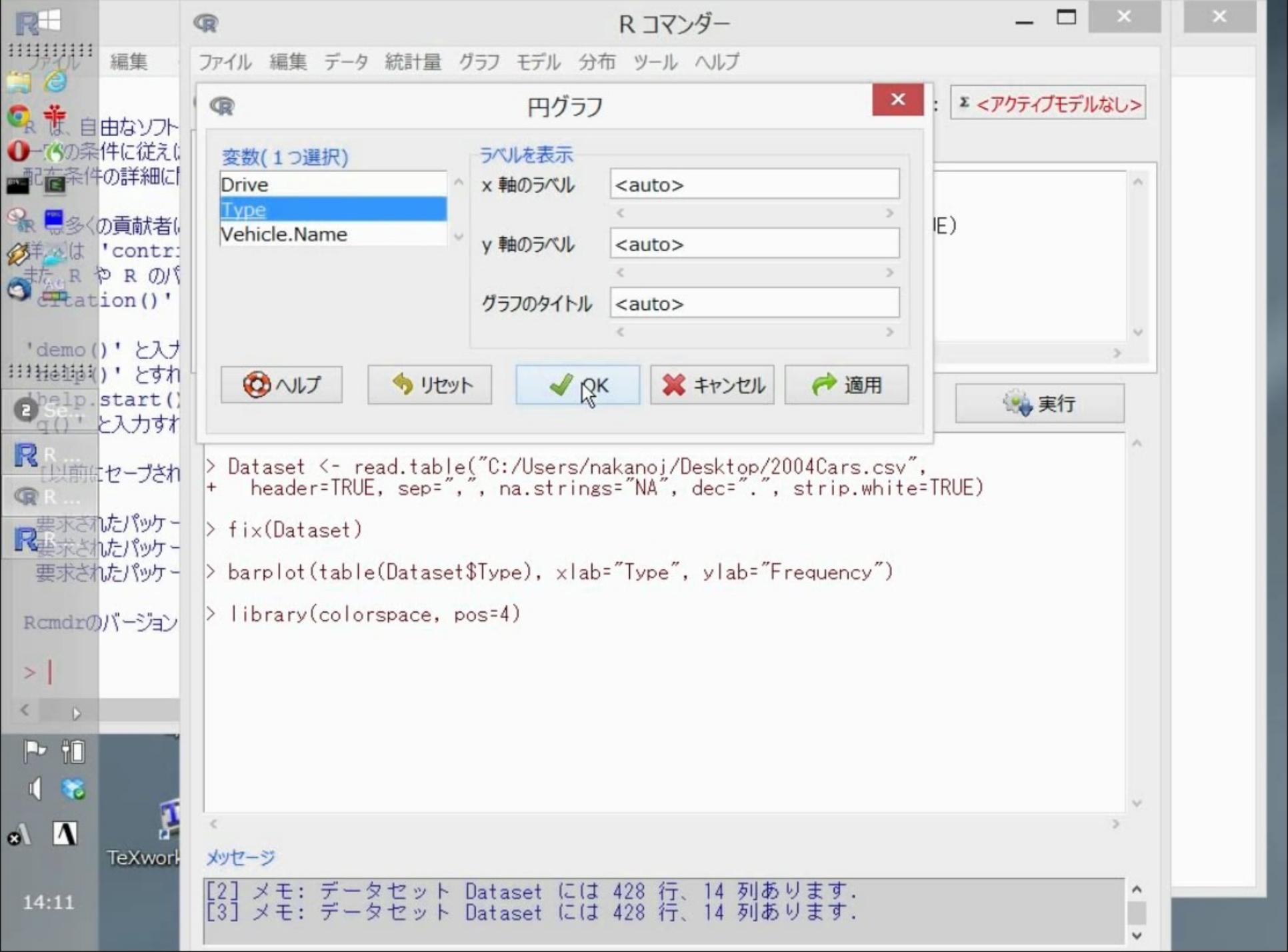
円グラフ...

3次元グラフ ▶

グラフをファイルに保存 ▶

メッセージ

```
[2] メモ: データセット Dataset には 428 行、14 列あります。
[3] メモ: データセット Dataset には 428 行、14 列あります。
```



円グラフ

変数 (1つ選択)

- Drive
- Type
- Vehicle.Name

ラベルを表示

x 軸のラベル <auto>

y 軸のラベル <auto>

グラフのタイトル <auto>

ヘルプ

リセット

OK

キャンセル

適用

<アクティブモデルなし>

実行

```
> Dataset <- read.table("C:/Users/nakanoj/Desktop/2004Cars.csv",
+ header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
> fix(Dataset)
> barplot(table(Dataset$Type), xlab="Type", ylab="Frequency")
> library(colorspace, pos=4)
```

メッセージ

[2] メモ: データセット Dataset には 428 行、14 列あります。
 [3] メモ: データセット Dataset には 428 行、14 列あります。

Windows taskbar and desktop icons including Start menu, taskbar, and system tray.

R GUI menu and toolbar with options like ファイル (File), 編集 (Edit), データ (Data), 統計 (Statistics), データセット (Dataset), and Rスクリプト (R Script).

R Graphics: Device 2 (ACTIVE)

ファイル 履歴 サイズ変更

Type

A pie chart titled 'Type' showing the frequency of different car types. The y-axis is labeled 'Frequency'. The chart is divided into six segments: Sedan (green, largest), Sports Car (teal), SUV (blue), Wagon (purple), Mini Van (pink), and Pickup (tan, smallest).

Type	Frequency
Sedan	~45
Sports Car	~15
SUV	~12
Wagon	~8
Mini Van	~5
Pickup	~3

出力

```
> Dataset <- read.table(
+ header=TRUE, sep="
> fix(Dataset)
> barplot(table(Dataset$Type),
+ library(
+ pie(table(Dataset$Type),
+ ylab="Frequency",
```

メッセージ

```
[2] メモ: データセット Dataset には 428 行、14 列あります。
[3] メモ: データセット Dataset には 428 行、14 列あります。
```

ファイル 編集 データ 統計量 **グラフ** モデル 分布 ツール ヘルプ

データセット: Data

データセットを表示

モデル: <アクティブモデルなし>

Rスクリプト Rマークダウン

```
Dataset <- read.table(
  header=TRUE, sep=",",
  fix(Dataset)
barplot(table(Dataset$
library(colorspace, po
pie(table(Dataset$Type
  ylab="Frequency", ma
```

色パレット...

インデックスプロット...

ヒストグラム...

密度推定...

幹葉表示...

箱ひげ図...

QQプロット...

散布図...

散布図行列...

折れ線グラフ...

条件付き散布図

平均のプロット...

ドットチャート...

棒グラフ...

円グラフ...

3次元グラフ ▶

グラフをファイルに保存 ▶

出力

```
> Dataset <- read.table(
+   header=TRUE, sep=",",
  Desktop/2004Cars.csv",
  dec=".", strip.white=TRUE)
  ylab="Frequency")
dataset$Type), xlab="Type",
  row_hcl(length(levels(Dataset$Type))))

> Dataset <- read.table(
+   header=TRUE, sep=",",
  Desktop/2004Cars.csv",
  dec=".", strip.white=TRUE)
  ylab="Frequency")
  labels=levels(Dataset$Type), xlab="Type",
+   ylab="Frequency", main="Type", col=rainbow_hcl(length(levels(Dataset$Type))))
```

実行

メッセージ

```
[2] メモ: データセット Dataset には 428 行、14 列あります。
[3] メモ: データセット Dataset には 428 行、14 列あります。
```

散布図

データ オプション

x 変数 (1つ選択)

City.Miles.Per.Gallon
Dealer.Cost..USD.
Engine.Size..liters.
Highway.Miles.Per.Gallon
Horsepower
Length

y 変数 (1つ選択)

City.Miles.Per.Gallon
Dealer.Cost..USD.
Engine.Size..liters.
Highway.Miles.Per.Gallon
Horsepower
Length

層別のプロット...

部分集合の表現

<全ての有効なケース>

ヘルプ

リセット

OK

キャンセル

適用

メッセージ

[2] メモ: データセット Dataset には 428 行、14 列あります。
[3] メモ: データセット Dataset には 428 行、14 列あります。

Windows taskbar and desktop icons including R logo, TeXwork, and system tray icons.

散布図

データ オプション

プロットのオプション

- x の値にゆらぎを与えて表示
- y の値にゆらぎを与えて表示
- x 軸を対数軸に
- y 軸を対数軸に
- 周辺箱ひげ図
- 最小 2 乗直線
- 平滑線
- ばらつき幅の表示

スムージングの幅

点を特定

- 自動的に
- マウスでインタラクティブに
- 確認しない

確認する点の数

ラベルと点を表示

x 軸のラベル

y 軸のラベル

グラフのタイトル

プロットする記号

点の大きさ

軸テキストの大きさ

軸ラベルのテキストの大きさ

ヘルプ リセット OK キャンセル 適用

メッセージ

[2] メモ: データセット Dataset には 428 行、14 列あります。
 [3] メモ: データセット Dataset には 428 行、14 列あります。

Windows taskbar and sidebar icons including Start menu, taskbar, and system tray.

R Rcmdr interface showing menu options (ファイル, 編集, データ, 統計) and a script editor with R code. The code includes commands for reading data, fixing it, and plotting barplots and scatterplots. The output window shows the results of these commands, including the number of rows and columns in the dataset.

```

ファイル 編集 データ 統計
データセット: Dataset
Rスクリプト Rマークダウン
barplot(table(Dataset$Type))
library(colorspace)
pie(table(Dataset$Type))
ylab="Frequency",
scatterplot(Highway.Miles.Per.Gallon ~ City.Miles.Per.Gallon,
smooth=FALSE, span=0.5, data=Dataset)

```

出力

```

> Dataset <- read.table("data.csv",
+ header=TRUE, sep=";", as.is=TRUE)
> fix(Dataset)
> barplot(table(Dataset$Type))
> library(colorspace)
> pie(table(Dataset$Type))
+ ylab="Frequency"
> scatterplot(Highway.Miles.Per.Gallon ~ City.Miles.Per.Gallon,
+ smooth=FALSE, span=0.5, data=Dataset)
70 94
67 88

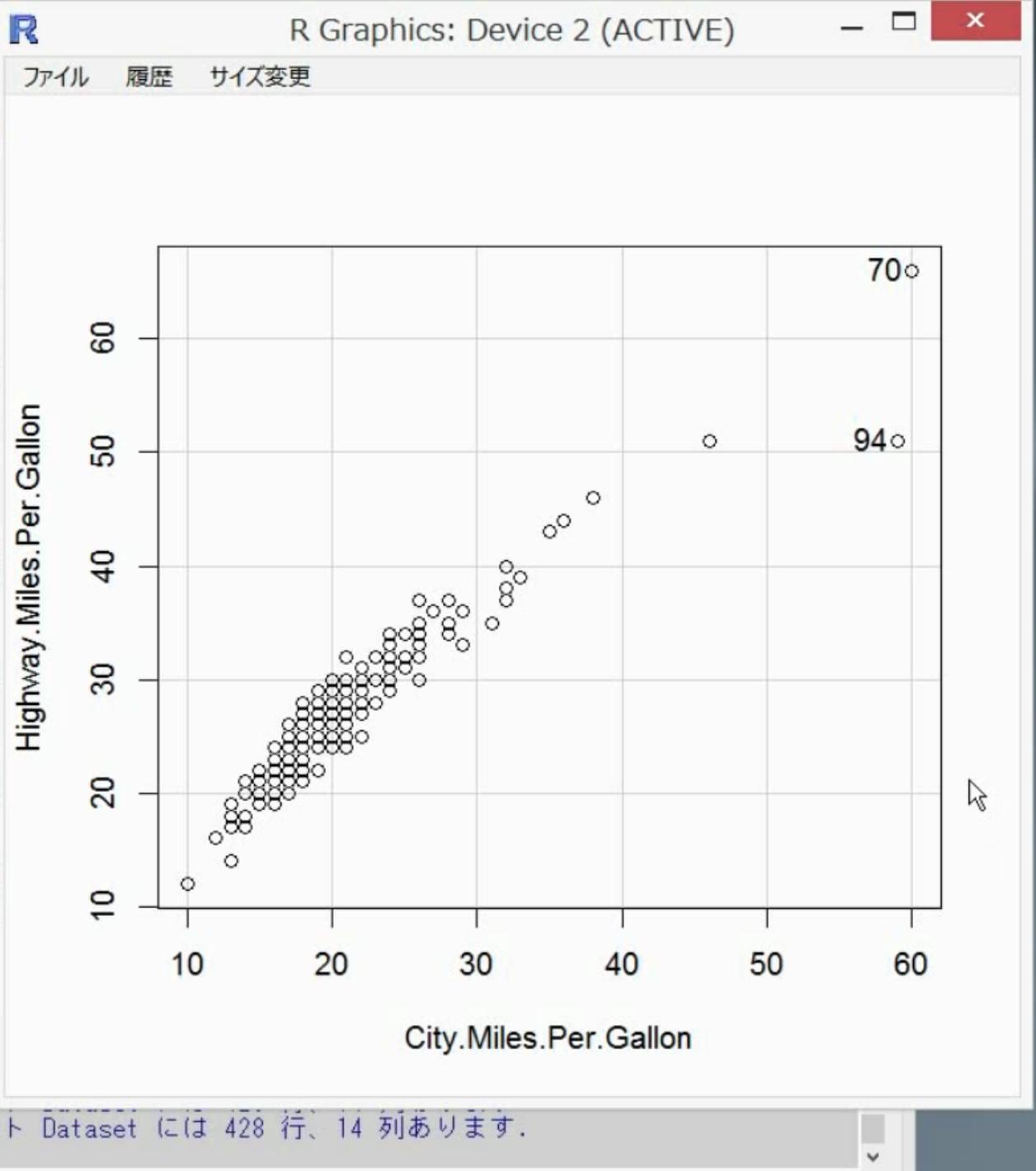
```

メッセージ

```

[2] メモ: データセット Dataset には 428 行、14 列あります。
[3] メモ: データセット Dataset には 428 行、14 列あります。

```



統計ソフトウェアRのグラフィックス例

- どのようなグラフィックスがRで描けるかは
“R Graphical Manual”
<http://rgm3.lab.nig.ac.jp/RGM/> を見ると良い
 - Rのパッケージの(ほとんど)すべてのグラフィックス(38000枚以上)がWeb上で閲覧でき、その作成プログラムを見ることができる

R Graphical Manual

search

Source Code Search Blog (test) H

Browse All

Last data update: 2013.05.20

Results 2945 - 2960 of 38596 found.

[1] < 180 181 182 183 184 185 186 187 188 189 190 > [2413] 185 Sort: -

Data Source

- R Release (2.15.3)
- CranContrib
- BioConductor
- All

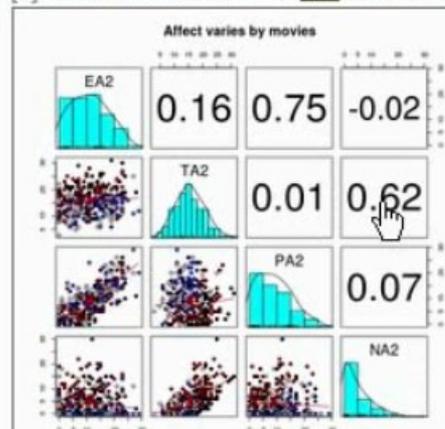
Data Type

- Packages
- Functions
- Images
- Data set

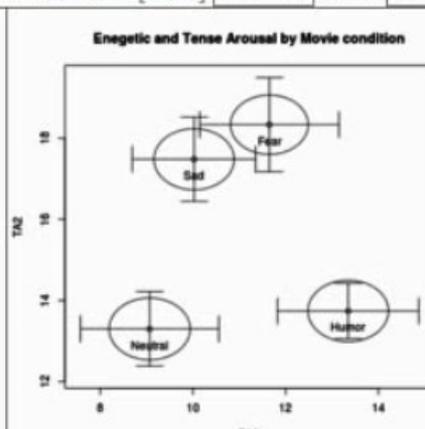
Classification

Cran Task View

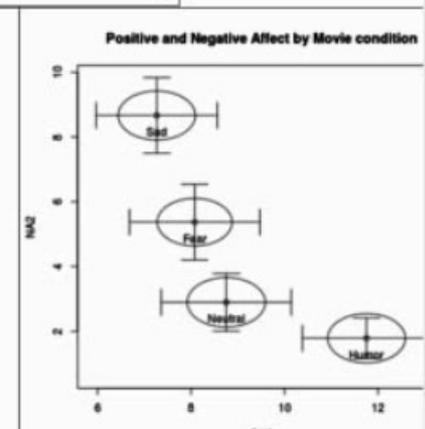
- Bayesian (810)
- ChemPhys (738)
- ClinicalTrials (466)
- Cluster (602)
- DifferentialEquations (549)
- Distributions (1319)
- Econometrics (1200)
- Environmetrics (2620)
- ExperimentalDesign (309)
- Finance (838)



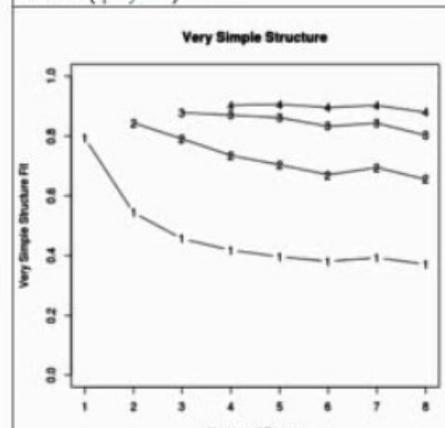
affect (psych)



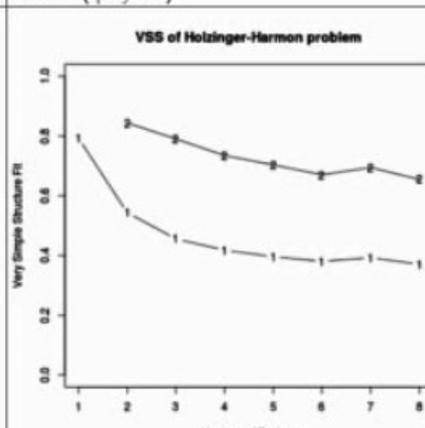
affect (psych)



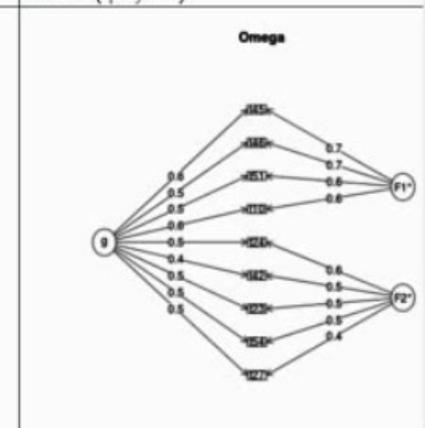
affect (psych)



(psych)

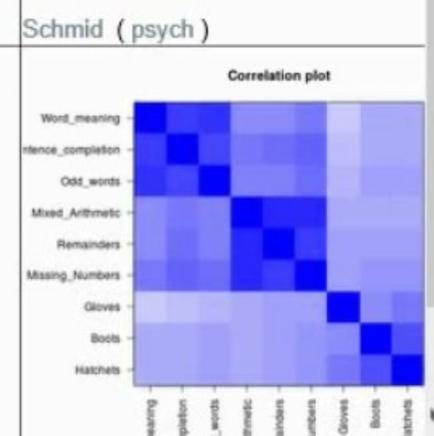
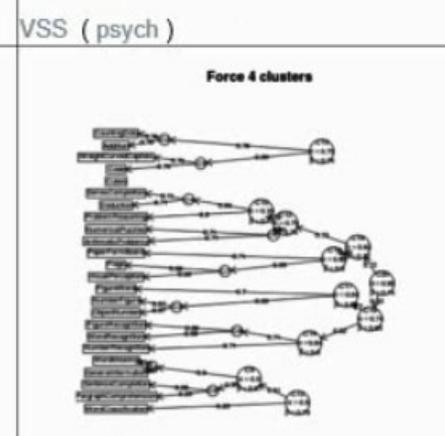
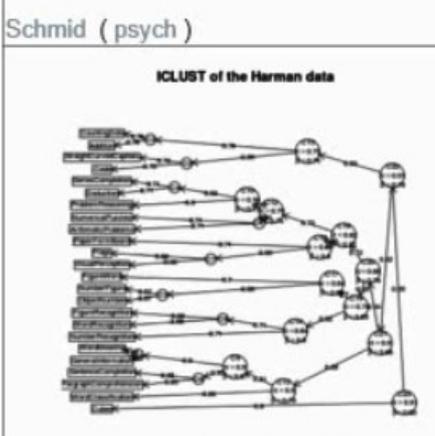
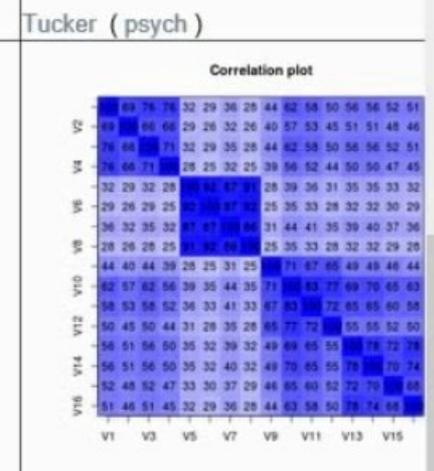
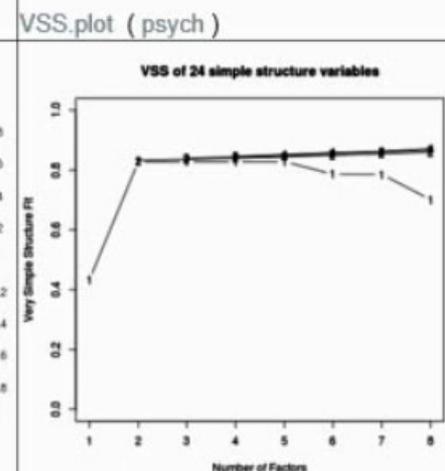
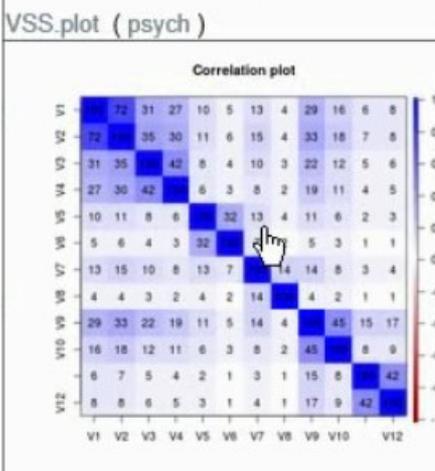
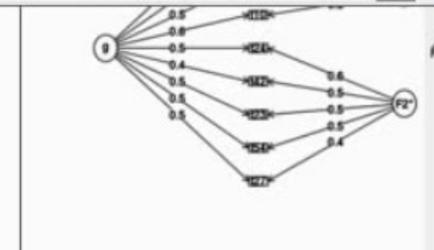
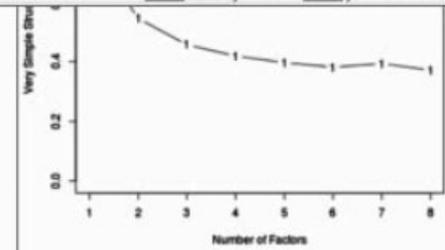
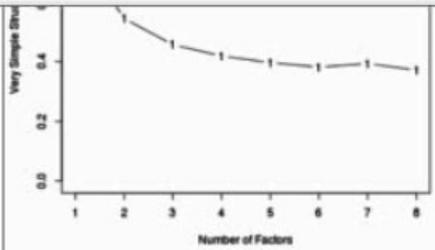


(psych)



Tucker (psych)

- ClinicalTrials (466)
- Cluster (602)
- DifferentialEquations (549)
- Distributions (1319)
- Econometrics (1200)
- Environmetrics (2620)
- ExperimentalDesign (309)
- Finance (838)
- Genetics (402)
- Graphics (510)
- HighPerformanceComputing (87)
- MachineLearning (604)
- MedicalImaging (67)
- MetaAnalysis (57)
- Multivariate (1571)
- NaturalLanguageProcessing (98)
- OfficialStatistics (468)
- Optimization (190)
- Pharmacokinetics (28)
- Phylogenetics (756)
- Psychometrics (321)
- ReproducibleResearch (0)
- Robust (240)
- SocialSciences (907)
- Spatial (2915)
- SpatioTemporal (852)
- Survival (1851)
- TimeSeries (1388)
- gR (102)
- Other (23057)





x

前方

広辞苑 第六版 DVD-ROM版 - 動画・画



R Graphical Manual

search

[Source Code Search](#) [Blog \(test\)](#) [Help \(test\)](#)[Browse All](#)

Last data update: 2013.05.20

Schmid

R Documentation

Data Source

- R Release (2.15.3)
- CranContrib
- BioConductor
- All

Data Type

- Packages
- Functions
- Images
- Data set

12 variables created by Schmid and Leiman to show the Schmid-Leiman Transformation

Description

John Schmid and John M. Leiman (1957) discuss how to transform a hierarchical factor structure to a bifactor structure. Schmid contains the example 12 x 12 correlation matrix. `schmid.leiman` is a 12 x 12 correlation matrix with communalities on the diagonal. This can be used to show the effect of correcting for attenuation. Two additional data sets are taken from Chen et al. (2006).

Usage

```
data(Schmid)
```

Details

Two artificial correlation matrices from Schmid and Leiman (1957). One real and one artificial covariance matrices from Chen et al. (2006).

- Schmid: a 12 x 12 artificial correlation matrix created to show the Schmid-Leiman transformation.
- `schmid.leiman`: A 12 x 12 matrix with communalities on the diagonal. Treating this as a covariance matrix shows the 6 x 6 factor solution
- Chen: An 18 x 18 covariance matrix of health related quality of life items from Chen et al. (2006). Number of observations = 403. The first item is a measure of the quality of life. The remaining 17 items form four subfactors: The items are (a) Cognition subscale: "Have difficulty reasoning and solving problems?" "React slowly to things that were said or done?"; "Become confused and start several actions at a time?" "Forget where you put things or appointments?"; "Have difficulty concentrating?" (b) Vitality subscale: "Feel tired?" "Have enough energy to do the things you want?" (R) "Feel worn out?"; "Feel full of pep?" (R). (c) Mental health subscale: "Feel calm and peaceful?" (R) "Feel downhearted and blue?"; "Feel very happy" (R); "Feel very nervous?"; "Feel so down in the dumps nothing could cheer you up?" (d) Disease worry subscale: "Were you afraid because of your health?"; "Were you frustrated about your health?"; "Was your health a worry in your life?"

3. 動的なグラフィックス

(代表的な)動的なグラフィックス

- **3次元散布図**

- 三つの連続変数データの分布
- 2次元平面上の(3次元)散布図を時間的に変化させることにより立体感を出す

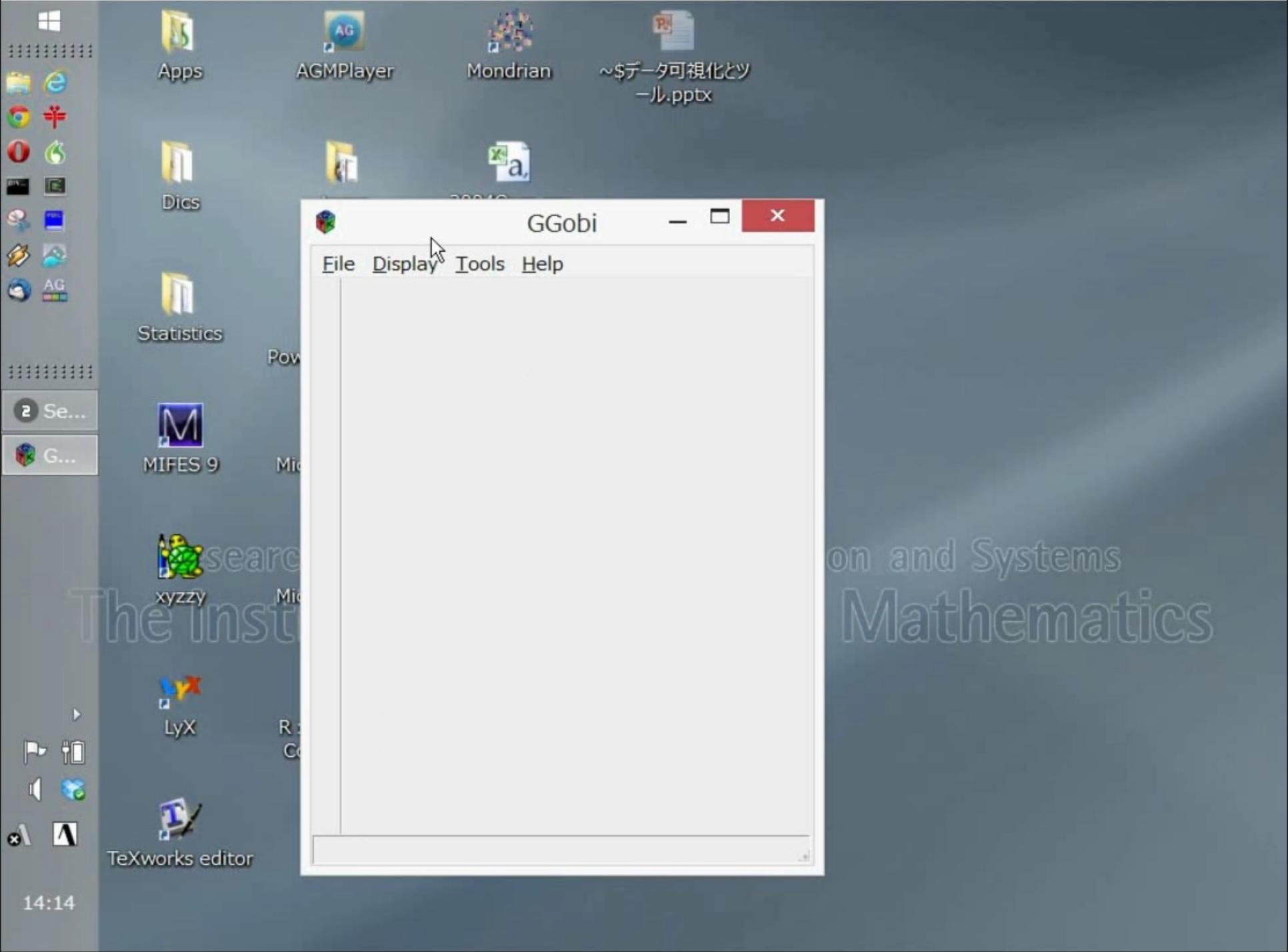
- **射影追跡**

- 三つ以上の連続変数データの分布
- 多次元データを、時間的に変化する直線または2次元平面に射影し、その表示により多次元データを多方向から眺める

動的なグラフィックスのためのソフトウェア

- 動的なグラフィックス機能を備えた(フリー)ソフトウェアとしては GGobi が有名
 - <http://www.ggobi.org/>
 - “GGobi is an open source visualization program for exploring high-dimensional data”
- Ggobiには対話的な機能も豊富
- 機能豊富なためもあり、操作が若干複雑





Apps

AGMPlayer

Mondrian

~\$データ可視化とツ
ール.pptx

Dics

Statistics

MIFES 9

xyzy

LyX

TeXworks editor

GGobi

File Display Tools Help

14:14

on and Systems
Mathematics

Read ggobi data

nakanoj Desktop

- Places
- Search
- Recently Used
- nakanoj
- Desktop
- Windows 8 (C:)
- SD (D:)

Name	Size	Modified
Statistics		Yesterday at 18:49
temp		月曜日
データ可視化とツール		Yesterday at 20:09
2004Cars.csv	33.1 KB	月曜日
AGDRec.agm	267.7 MB	14:03
AGMPlayer.lnk	3.0 KB	14:00
ggobi.lnk	1.6 KB	月曜日
LyX.lnk	2.0 KB	月曜日
Microsoft Excel 2010.lnk	2.7 KB	2013/02/04
Microsoft PowerPoint 2010.lnk	2.7 KB	月曜日
Microsoft Word 2010.lnk	2.7 KB	2013/02/03
Mondrian.lnk	636 bytes	Yesterday at 19:29
R x64 3.0.2 (R Commander).lnk	1.2 KB	Yesterday at 18:50
TeXworks editor.lnk	887 bytes	2013/11/15
Titanic.txt	44.0 KB	2013/09/17

Add Remove

Input Type: unknown URL:

Cancel Open



GGobi

File Display **View** Interaction Tools Help

XY Plot

Plot cycling

- Cycle

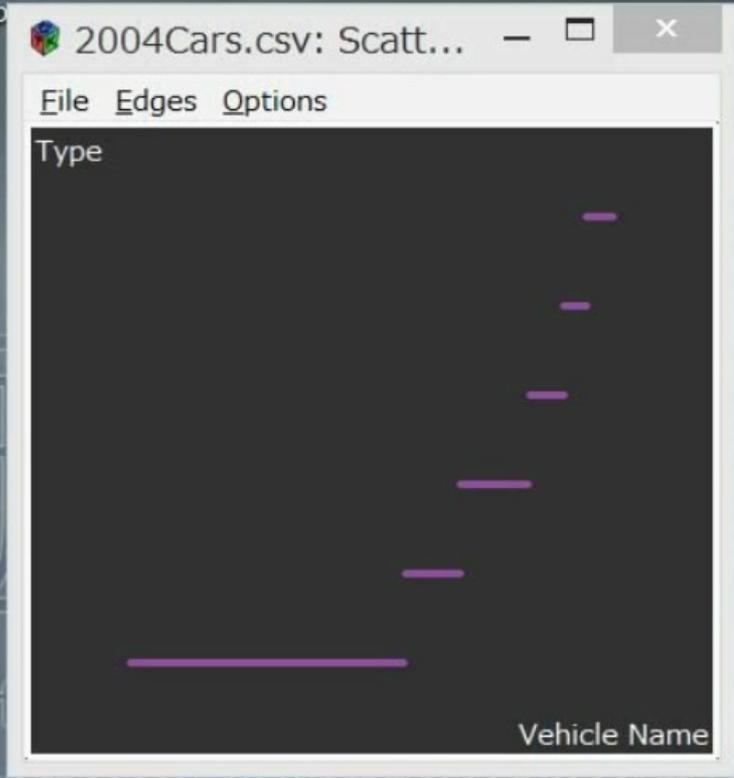
No fixed axes

Change direction

X	Y	Engine Size (liters)
X	Y	Number of Cylinders
X	Y	Horsepower
X	Y	City Miles Per Gallon
X	Y	Highway Miles Per Gallon
X	Y	Weight (Pounds)
X	Y	Wheel Base (inches)
X	Y	Length
X	Y	Width (inches)

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\...)

- 1D Plot Ctrl+D
- XY Plot Ctrl+X
- 1D Tour Ctrl+T
- Rotation Ctrl+R**
- 2D Tour Ctrl+G
- 2x1D Tour Ctrl+U



LyX R x64 3.0.2 (R データ可視化とツール Commander)



Windows taskbar icons including File Explorer, Internet Explorer, Google Chrome, Firefox, and various utility icons.

GGobi

File Display View Interaction Tools Help

Rotation:

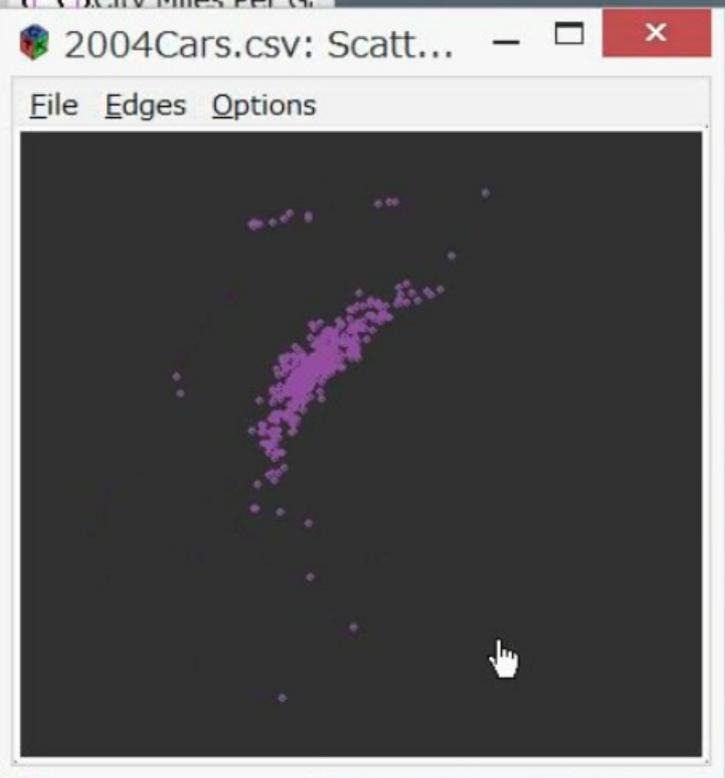
Pause

Reinit Scramble

Manual manipulation: Oblique

X	Y	Z	Vehicle N
X	Y	Z	Type
X	Y	Z	Drive
X	Y	Z	Suggester
X	Y	Z	Dealer Co
X	Y	Z	Engine Si
X	Y	Z	Number c
X	Y	Z	Horsepow
X	Y	Z	City Miles
X	Y	Z	Highway I
X	Y	Z	Weight (F
X	Y	Z	Wheel Ba
X	Y	Z	Length
X	Y	Z	Width (in

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\2004Cars.csv)



System tray icons including volume, network, and system clock.

Taskbar showing active applications: LyX, R x64 3.0.2 (R データ可視化とツール Commander), TeXworks editor, ggobi, and AGDRec.agm.

14:14

Windows taskbar icons including Start button, File Explorer, Internet Explorer, Google Chrome, Firefox, VLC media player, and various application shortcuts.

GGobi

File Display **View** Interaction Tools Help

Rotation

Pause

Reinit

Manual manipulation

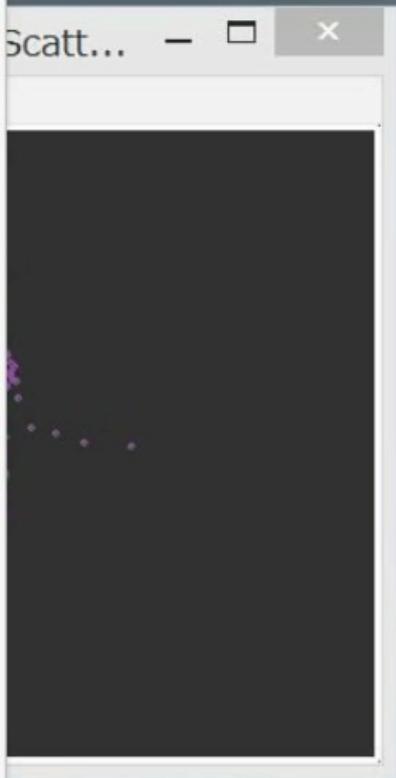
Oblique

X	Y	Z	Vehicle N
X	Y	Z	Type
X	Y	Z	Drive
X	Y	Z	Suggested
X	Y	Z	Dealer Co
X	Y	Z	Engine Si
X	Y	Z	Number c
X	Y	Z	Horsepow
X	Y	Z	City Miles
X	Y	Z	Highway I
X	Y	Z	Weight (F
X	Y	Z	Wheel Ba
X	Y	Z	Length
X	Y	Z	Width (in

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\2004Cars.csv)

- 1D Not Ctrl+D
- XY Plot Ctrl+X
- 1D Tour Ctrl+T
- Rotation Ctrl+R
- 2D Tour Ctrl+G
- 2x1D Tour Ctrl+U

- City Miles Per Gallon
- Highway Miles Per Gallon
- Weight (Pounds)



LyX R x64 3.0.2 (R データ可視化とツール Commander)

Taskbar icons: TeXworks editor, ggobi, AGDRec.agm

Windows taskbar showing icons for File Explorer, Internet Explorer, Google Chrome, Firefox, VLC media player, and other applications. A search bar is visible with the text "2 Se...".

GGobi

File Display View Interaction Tools Help

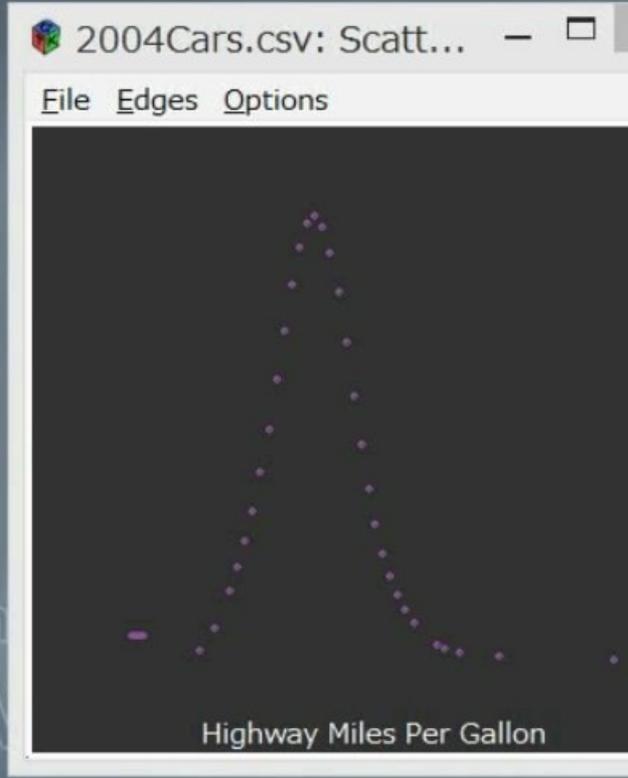
1D Plot
 ASH ▼

ASH parameters
 ASH: add lines
 ASH smoothness:
0.20

Plot cycling
 Cycle
Change direction

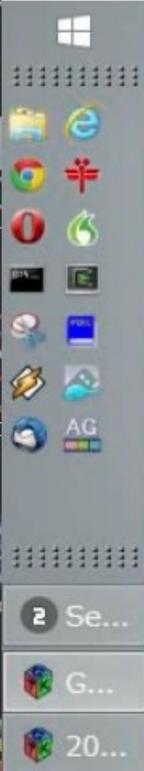
- Vehicle Name
- Type
- Drive
- Suggested Retail Price (USD)
- Dealer Cost (USD)
- Engine Size (liters)
- Number of Cylinders
- Horsepower
- City Miles Per Gallon
- Highway Miles Per Gallon
- Weight (Pounds)
- Wheel Base (inches)
- Length
- Width (inches)

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\2004Cars.csv)



System tray area showing icons for volume, network, and other background services. The time 14:15 is displayed at the bottom left.

Taskbar showing active applications: LyX, R x64 3.0.2 (R データ可視化とツール Commander), TeXworks editor, ggobi, and AGDRec.agm.



GGobi

File Display **View** Interaction Tools Help

1D Plot

ASH

ASH parameters

ASH: add

ASH smooth

0.20

Plot cycling

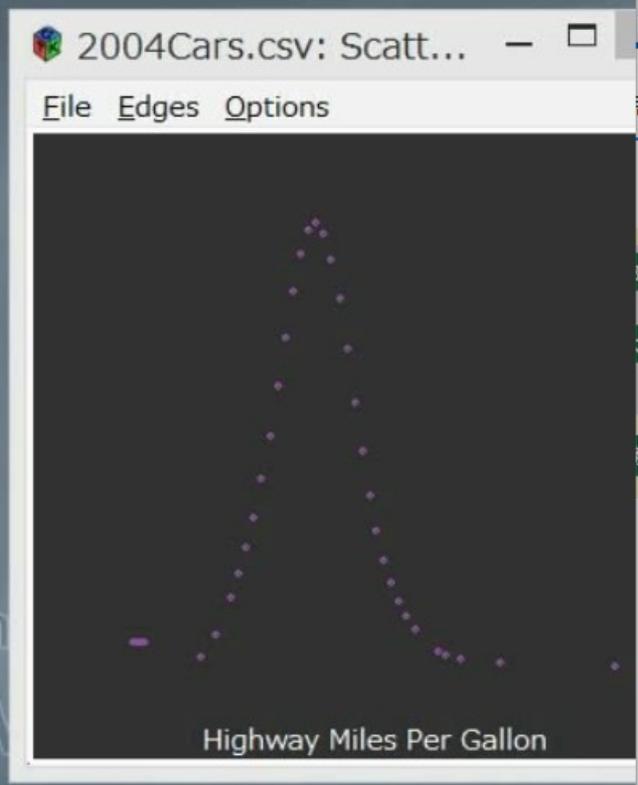
Cycle

Change direction

- 1D Plot Ctrl+D
- XY Plot Ctrl+X
- 1D Tour Ctrl+T
- Rotations Ctrl+R
- 2D Tour Ctrl+G
- 2x1D Tour Ctrl+U

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\2004Cars.csv)

<input type="checkbox"/>	Engine Size (liters)
<input type="checkbox"/>	Number of Cylinders
<input type="checkbox"/>	Horsepower
<input type="checkbox"/>	City Miles Per Gallon
<input checked="" type="checkbox"/>	Highway Miles Per Gallon
<input type="checkbox"/>	Weight (Pounds)
<input type="checkbox"/>	Wheel Base (inches)
<input type="checkbox"/>	Length
<input type="checkbox"/>	Width (inches)



LyX R x64 3.0.2 (R データ可視化とツール Commander)

TeXworks editor ggobi AGDRec.agm

GGobi

File Display View Interaction Tools Help

1D Tour

Pause

Reinit Scramble

ASH parameters

ASH: add lines

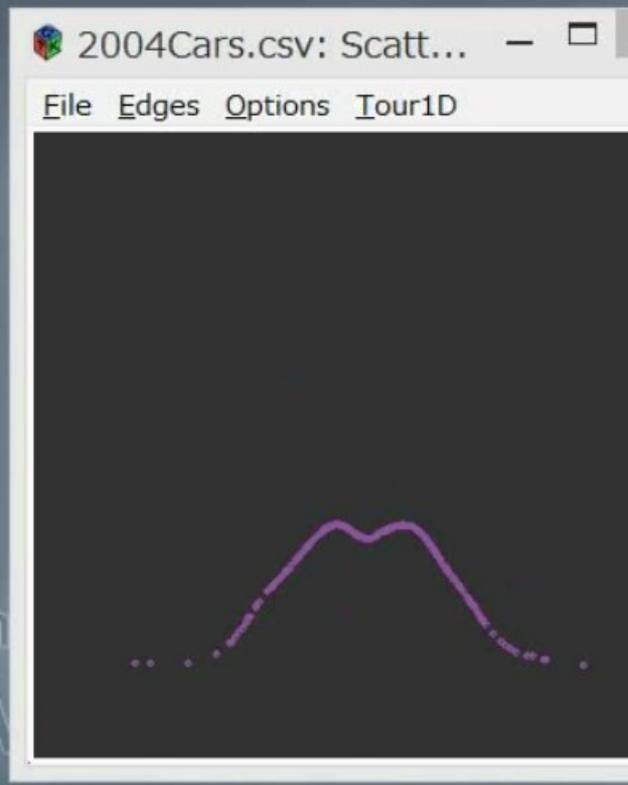
ASH smoothness: 0.19

Projection pursuit ...

<input checked="" type="checkbox"/>	Vehicle Name
<input checked="" type="checkbox"/>	Type
<input checked="" type="checkbox"/>	Drive
<input checked="" type="checkbox"/>	Suggested Retail Price
<input checked="" type="checkbox"/>	Dealer Cost (USD)
<input checked="" type="checkbox"/>	Engine Size (liters)
<input checked="" type="checkbox"/>	Number of Cylinder
<input checked="" type="checkbox"/>	Horsepower
<input checked="" type="checkbox"/>	City Miles Per Gallon
<input checked="" type="checkbox"/>	Highway Miles Per Gallon
<input checked="" type="checkbox"/>	Weight (Pounds)
<input checked="" type="checkbox"/>	Wheel Base (inches)
<input checked="" type="checkbox"/>	Length
<input checked="" type="checkbox"/>	Width (inches)

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\2004Cars.csv)

Manip



LyX R x64 3.0.2 (R データ可視化とツール Commander)

TeXworks editor ggobi AGDRec.agm

GGobi

File Display **View** Interaction Tools Help

1D Tour

Pause

Reinit

ASH parameters

ASH: add lines

ASH smoothness: 0.19

Projection pursuit ...

<input type="checkbox"/>	1D Plot	Ctrl+D
<input type="checkbox"/>	XY Plot	Ctrl+X
<input checked="" type="checkbox"/>	1D Tour	Ctrl+T
<input type="checkbox"/>	Rotation	Ctrl+R
<input checked="" type="checkbox"/>	2D Tour	Ctrl+G
<input type="checkbox"/>	2x1D Tour	Ctrl+U

Engine Size (liters)
 Number of Cylinder
 Horsepower
 City Miles Per Gallo
 Highway Miles Per C
 Weight (Pounds)
 Wheel Base (inches)
 Length
 Width (inches)

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\2004Cars.csv)

s.csv: Scatt... Options Tour1D

Horsepower
 City Miles Per G
 Highway Miles P
 Weight (Pounds
 Wheel Base (inc

Windows taskbar and desktop icons:

- Start menu
- Taskbar: Se..., G..., 20...
- Desktop icons: Apps, Dics, Statistics, MIFES 9, xyzy, LyX, TeXworks editor, ggobi, AGDRec.agm
- System tray: 14:17

R x64 3.0.2 (R データ可視化とツール Commander)

GGobi

File Display View Interaction Tools Help

Oblique

Scramble

Manual manipulation:

Oblique

Projection pursuit ...

2 Se...

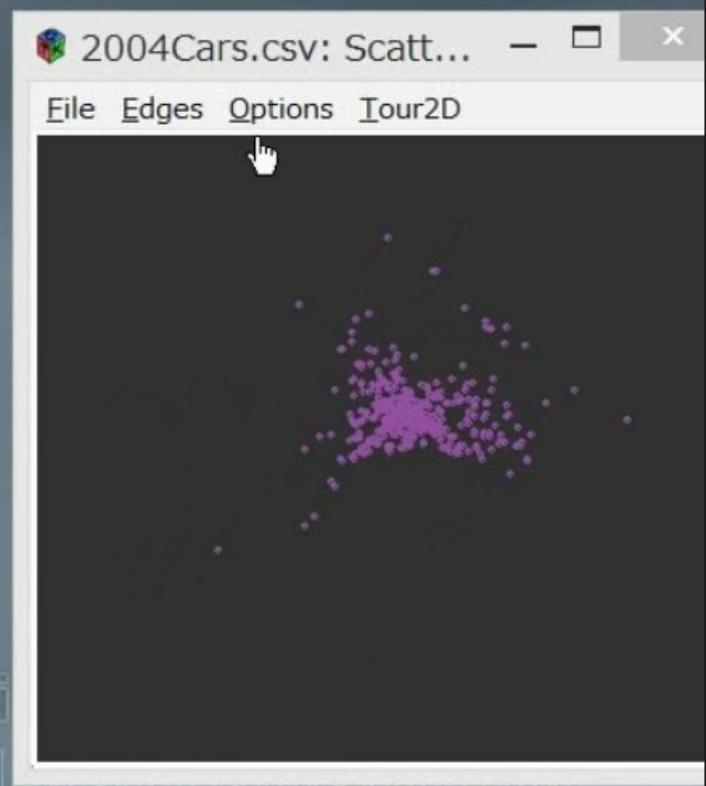
G...

20...

<input type="checkbox"/>	Vehicle Name		Horsepower
<input type="checkbox"/>	Type		City Miles Per Gallon
<input type="checkbox"/>	Drive		Highway Miles Per Gallon
<input type="checkbox"/>	Suggested Retail Price		Weight (Pounds)
<input type="checkbox"/>	Dealer Cost (USD)		Wheel Base (inches)
<input type="checkbox"/>	Engine Size (liters)		
<input type="checkbox"/>	Number of Cylinders		
<input checked="" type="checkbox"/>	Horsepower		
<input checked="" type="checkbox"/>	City Miles Per Gallon		
<input checked="" type="checkbox"/>	Highway Miles Per Gallon		
<input checked="" type="checkbox"/>	Weight (Pounds)		
<input checked="" type="checkbox"/>	Wheel Base (inches)		
<input type="checkbox"/>	Length		
<input type="checkbox"/>	Width (inches)		

Manip

2004Cars.csv: 428 x 14 (C:\Users\nakanoj\Desktop\2004Cars.csv)



Commander

TeXworks editor

ggobi

AGDRec.agm

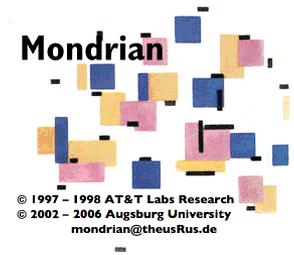
動的なグラフィックスに対する注意

- 動的なデータ可視化は物理現象の可視化ほど有用ではない
 - 物理現象は3次元のものが時間変化するので、そのためのグラフィックス(アニメーション)が非常に有用
 - 統計データは「多次元」であることが多いため、せいぜい3次元(および時間変化)を表すアニメーションでは不十分なことが多い
- 射影追跡はわかりやすいとはいえない
 - アニメーションを眺めて、有用な構造を発見する
 - そのアニメーションはランダムに動くことが多い
 - 発見もランダム＝重要な発見に出くわす確率は低い
- ただし、「人目を引く」効果は大きい

4. 対話的なグラフィックス

対話的なグラフィックスのためのソフトウェア

- フリーソフトウェア Mondrian が使い易い
 - “Mondrian - Interactive Statistical Data Visualization in JAVA”
<http://www.rosuda.org/mondrian/>
 - “Mondrian is a general purpose statistical data-visualization system”
- 対話的機能は、Rにはパッケージとして、GGobi にはもともとあるが、Mondrian のほうが**操作が容易**



対話的なグラフィックスの機能

- グラフィックス表示の**パラメータの指定**
 - 例1: ヒストグラムの階級幅、アンカー
 - 例2: **α チャンネル** (不透明度)
- 個体データの直接操作
 - **選択** (特定のデータを選ぶ)
 - **強調** (他のデータと色を変えて表示する)
 - **連携** (複数のグラフィックスでの共通の選択・強調など)
- 古典的なグラフであるヒストグラム、散布図でも有用

File Plot Calc Options Window Help

- Open Ctrl-O
- Open Recent
- Open R dataframe Ctrl+Shift-O
- Open Database Ctrl-D
- Save Ctrl-S
- Save Selection Ctrl+Shift-S
- Close Dataset Ctrl-W
- Quit Ctrl-Q

© 1997 - 1998 AT&T Labs Research
 © 2002 - 2006 Auesbure University
 Connection to R failed: Please check Rserve v1.2

Welcome!

mondrian

~\$データ可視化とツ
ール.pptx

04Cars.csv

tanic.txt

データ可視化とツ
ール.pptx

MIFES 9

Microsoft Excel 2010

データ可視化とツ
ール.pptx

xyzyzy

Microsoft Word 2010

ごみ箱

LyX

R x64 3.0.2 (R Commander)

データ可視化とツ
ール

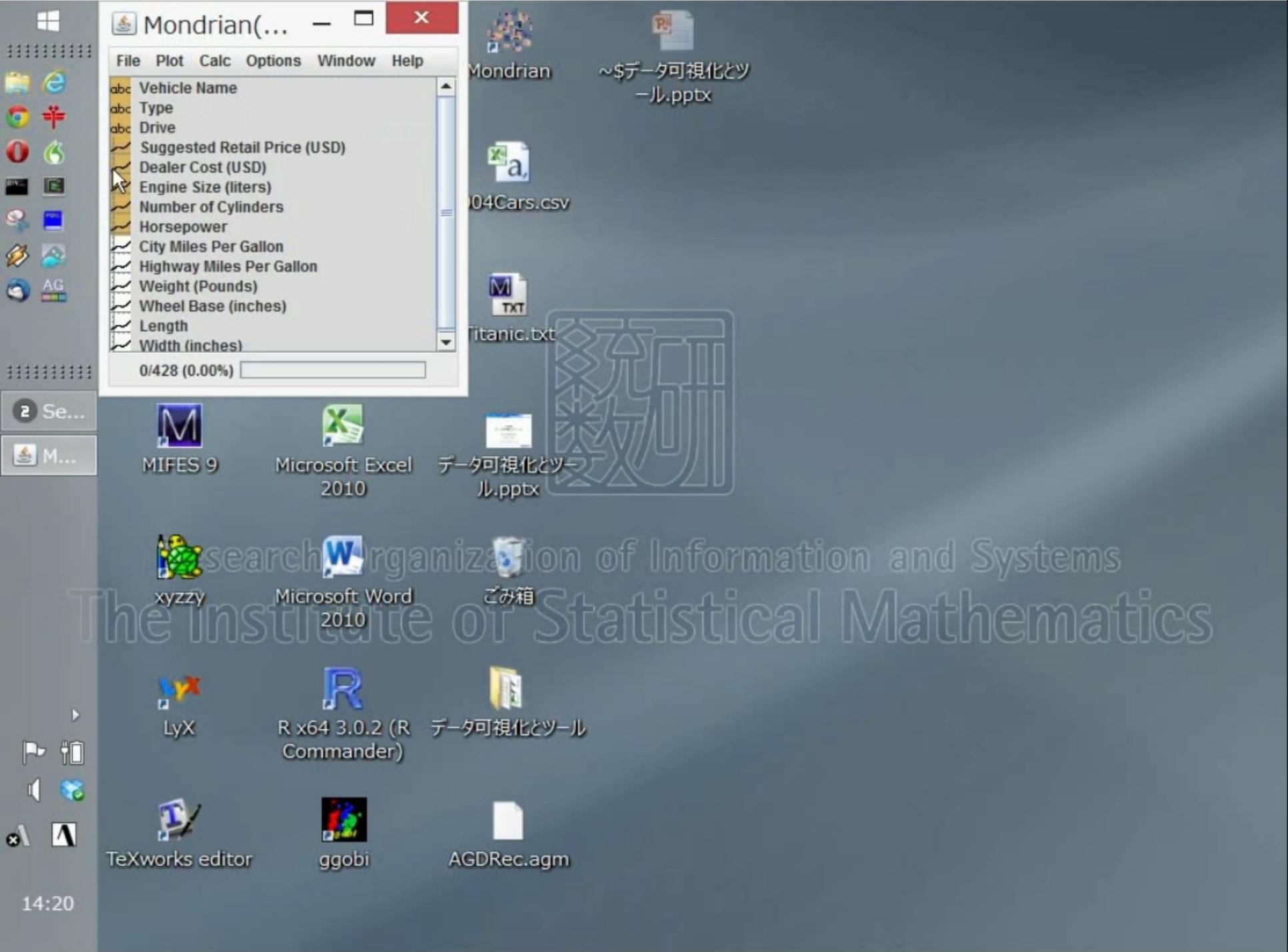
TeXworks editor

ggobi

AGDRec.agm



search organization of Information and Systems
 The Institute of Statistical Mathematics



Mondrian(...)

File Plot Calc Options Window Help

abc	Vehicle Name
abc	Type
abc	Drive
	Suggested Retail Price (USD)
	Dealer Cost (USD)
	Engine Size (liters)
	Number of Cylinders
	Horsepower
	City Miles Per Gallon
	Highway Miles Per Gallon
	Weight (Pounds)
	Wheel Base (inches)
	Length
	Width (inches)

0/428 (0.00%)

Mondrian

~\$データ可視化とツール.pptx

04Cars.csv

titanic.txt

MIFES 9

Microsoft Excel 2010

データ可視化とツール.pptx

xyzy

Microsoft Word 2010

ごみ箱

LyX

R x64 3.0.2 (R Commander)

データ可視化とツール

TeXworks editor

ggobi

AGDRec.agm



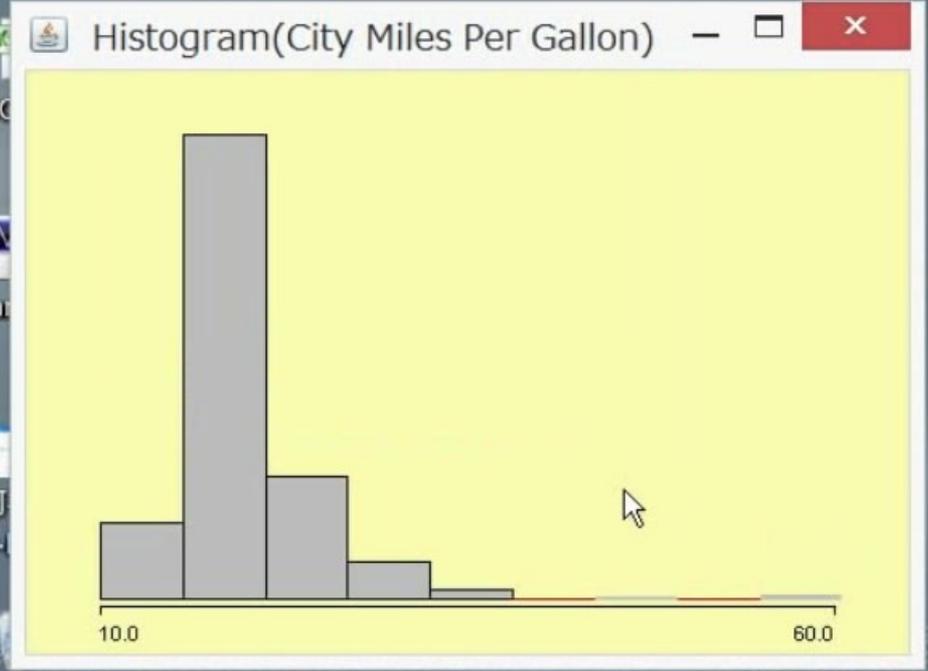
search organization of Information and Systems
The Institute of Statistical Mathematics

Mondrian(...)

File Plot Calc Options Window Help

- abc Vehicle Name
- abc Type
- abc Drive
- Suggested Retail Price (USD)
- Dealer Cost (USD)
- Engine Size (liters)
- Number of Cylinders
- Horsepower
- City Miles Per Gallon
- Highway Miles Per Gallon
- Weight (Pounds)
- Wheel Base (inches)
- Length
- Width (inches)

0/428 (0.00%)



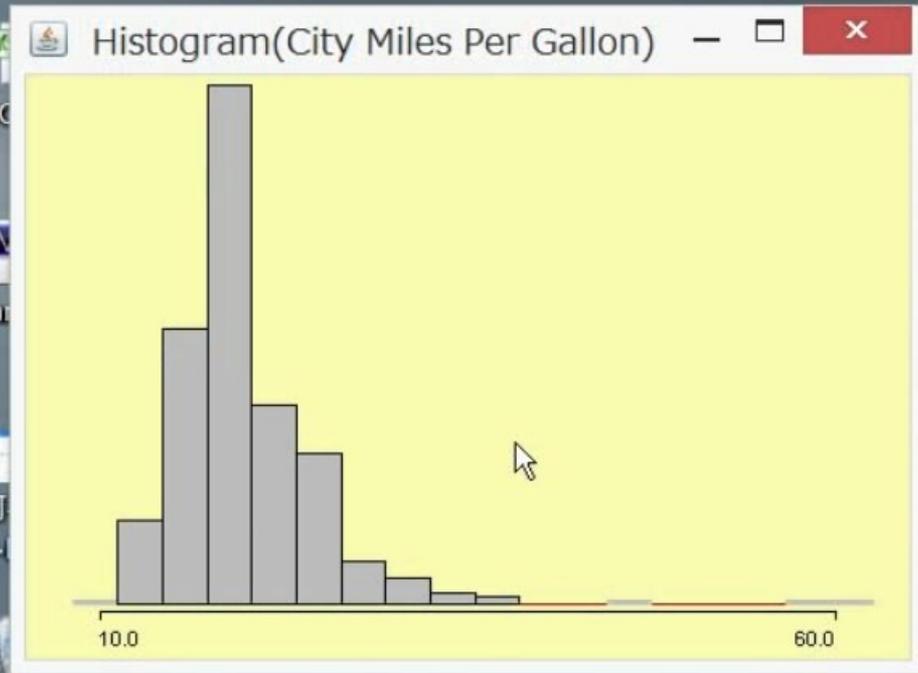
The Institute of Statistical Mathematics

Mondrian(...)

File Plot Calc Options Window Help

- abc Vehicle Name
- abc Type
- abc Drive
- Suggested Retail Price (USD)
- Dealer Cost (USD)
- Engine Size (liters)
- Number of Cylinders
- Horsepower
- City Miles Per Gallon
- Highway Miles Per Gallon
- Weight (Pounds)
- Wheel Base (inches)
- Length
- Width (inches)

0/428 (0.00%)



The Institute of Statistical Mathematics

Windows taskbar with icons for Start menu, Internet Explorer, Google Chrome, Firefox, VLC media player, Audacity, and other applications.

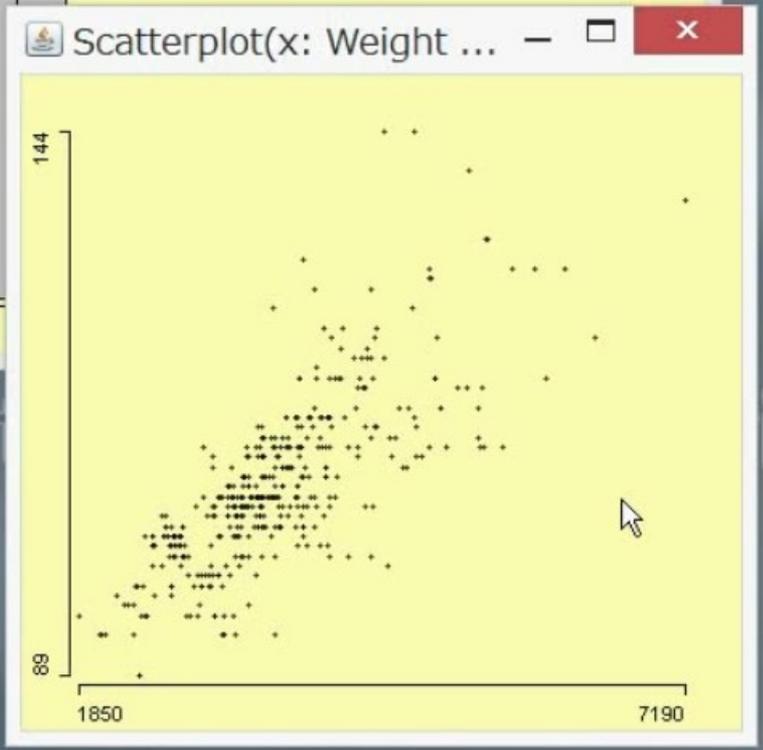
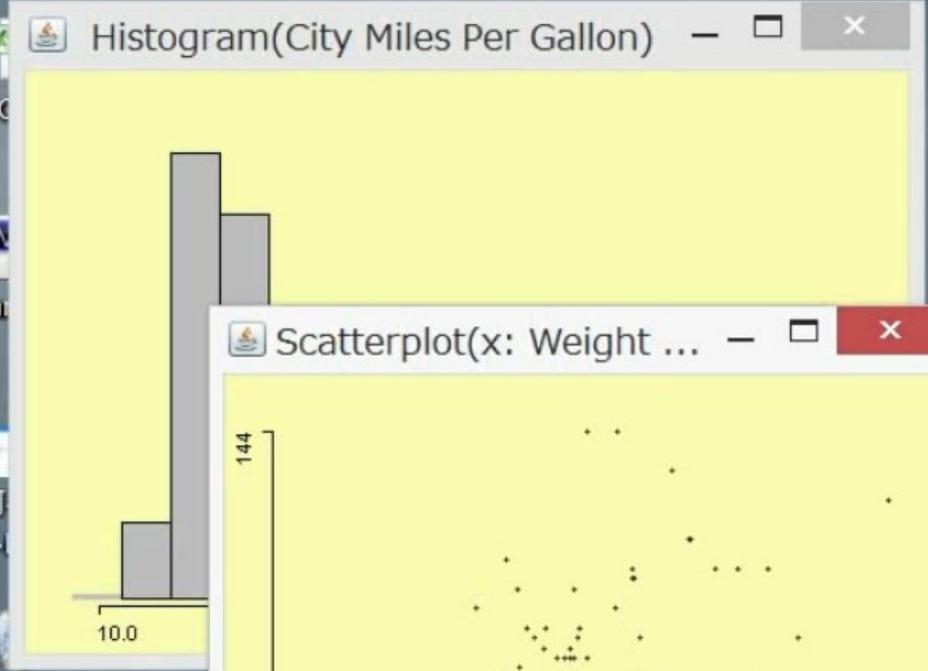
Mondrian(...)

File Plot Calc Options Window Help

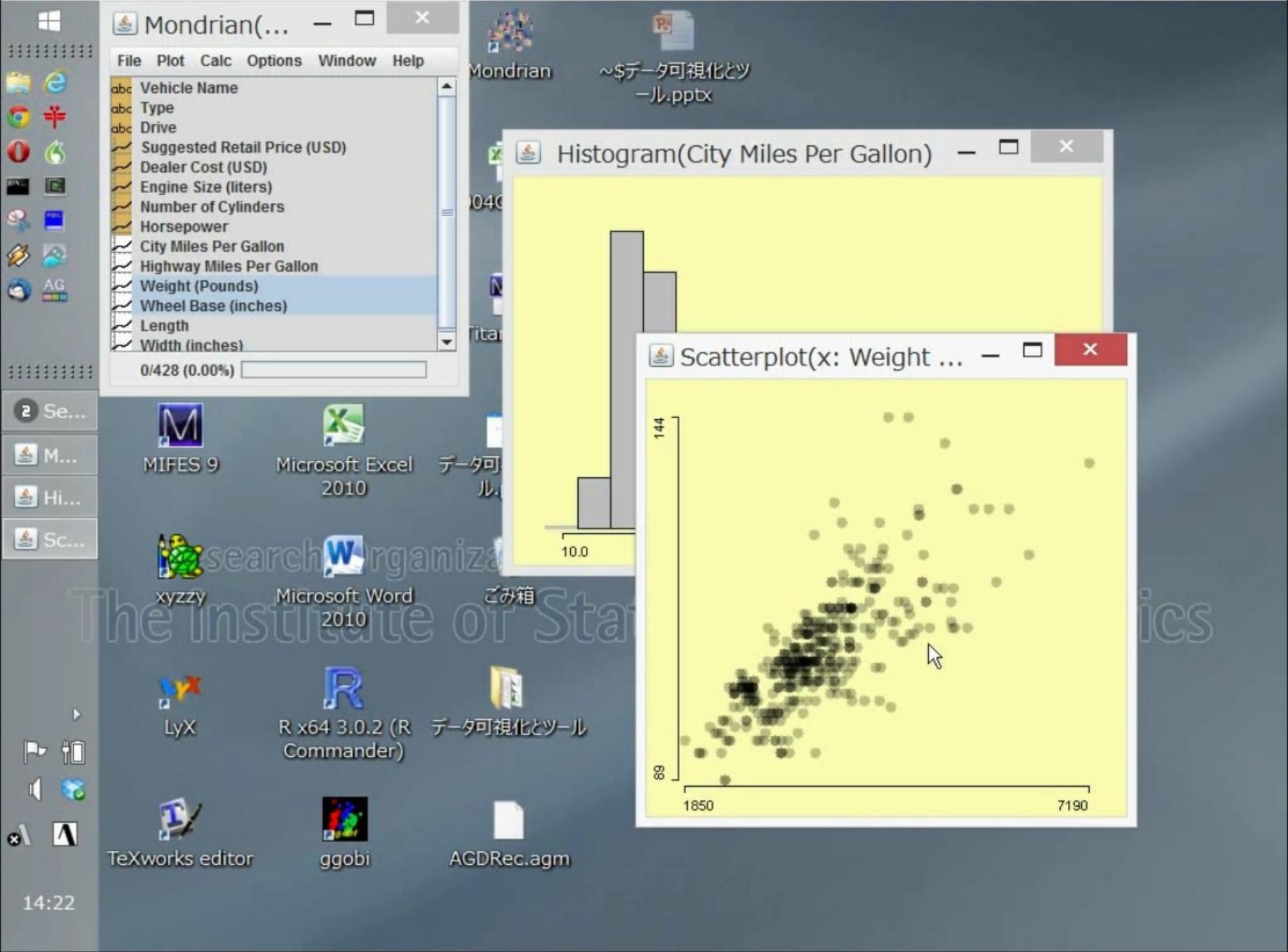
- abc Vehicle Name
- abc Type
- abc Drive
- Suggested Retail Price (USD)
- Dealer Cost (USD)
- Engine Size (liters)
- Number of Cylinders
- Horsepower
- City Miles Per Gallon
- Highway Miles Per Gallon
- Weight (Pounds)
- Wheel Base (inches)
- Length
- Width (inches)

0/428 (0.00%)

Mondrian ~\$データ可視化とツール.pptx



Desktop icons: MIFES 9, Microsoft Excel 2010, データ可視化とツール, xyzyzy, Microsoft Word 2010, ごみ箱, LyX, R x64 3.0.2 (R Commander), データ可視化とツール, TeXworks editor, ggobi, AGDRec.agm

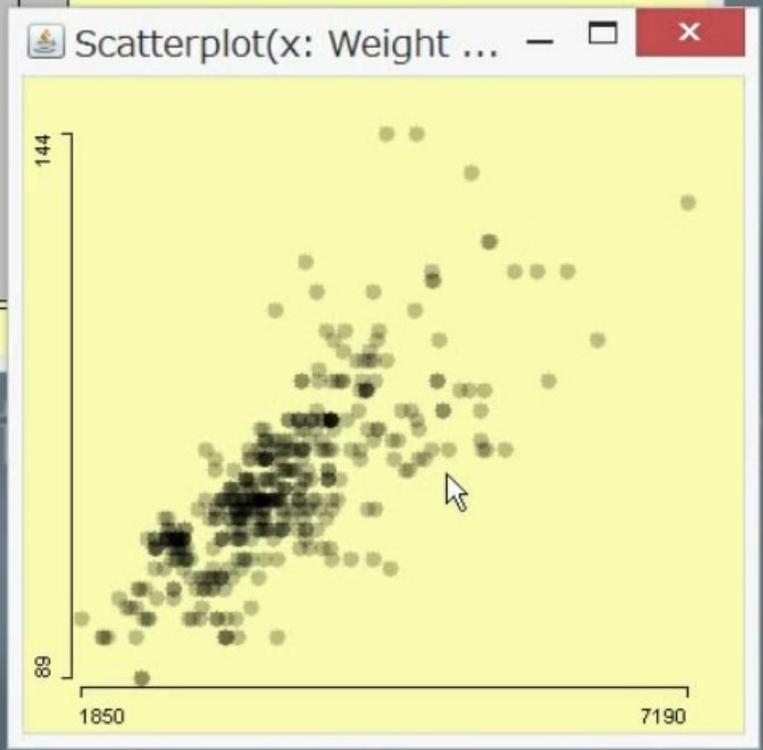
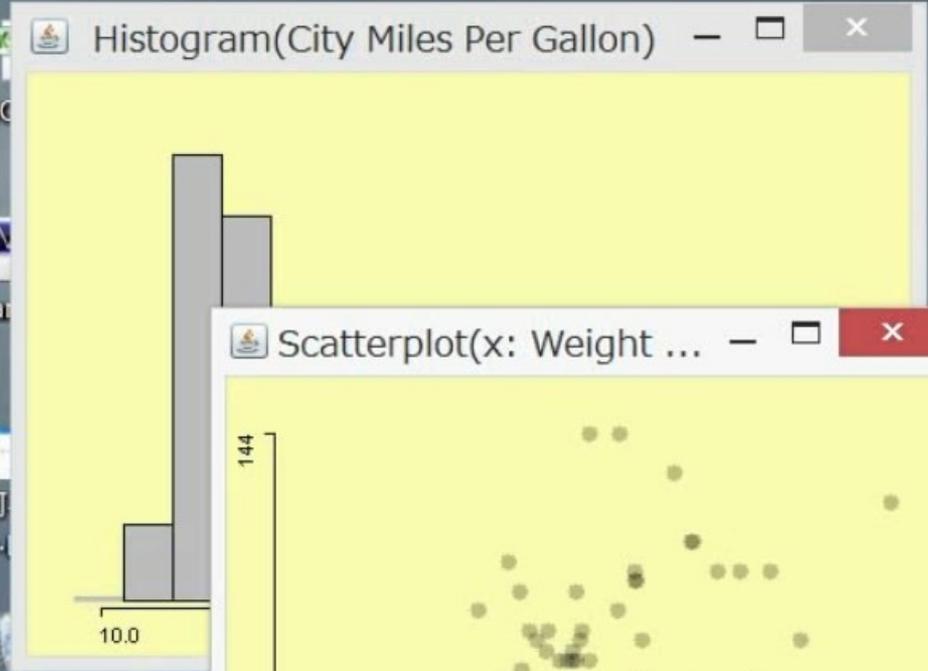


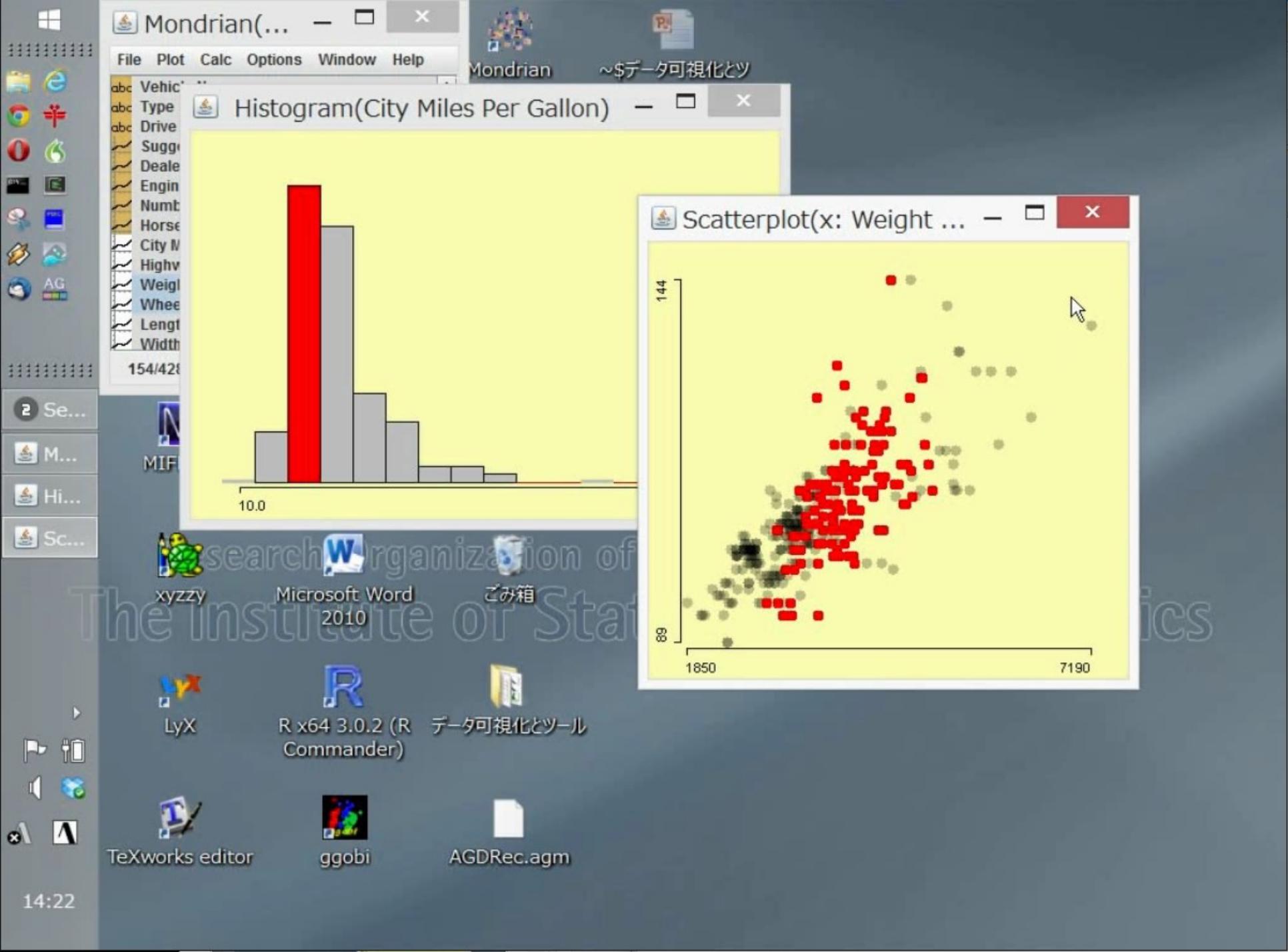
Mondrian(...)

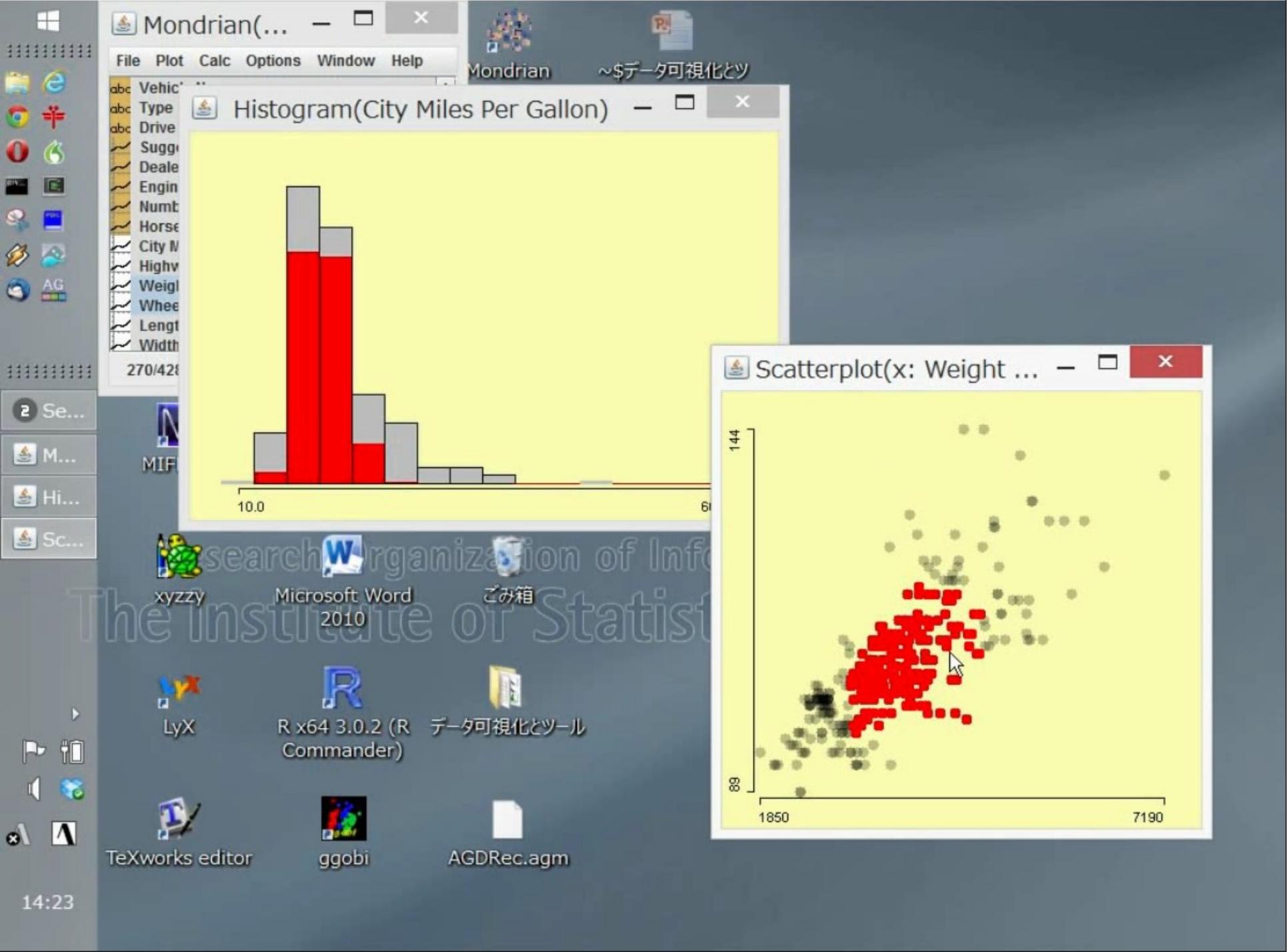
File Plot Calc Options Window Help

- abc Vehicle Name
- abc Type
- abc Drive
- ~ Suggested Retail Price (USD)
- ~ Dealer Cost (USD)
- ~ Engine Size (liters)
- ~ Number of Cylinders
- ~ Horsepower
- ~ City Miles Per Gallon
- ~ Highway Miles Per Gallon
- ~ Weight (Pounds)
- ~ Wheel Base (inches)
- ~ Length
- ~ Width (inches)

0/428 (0.00%)





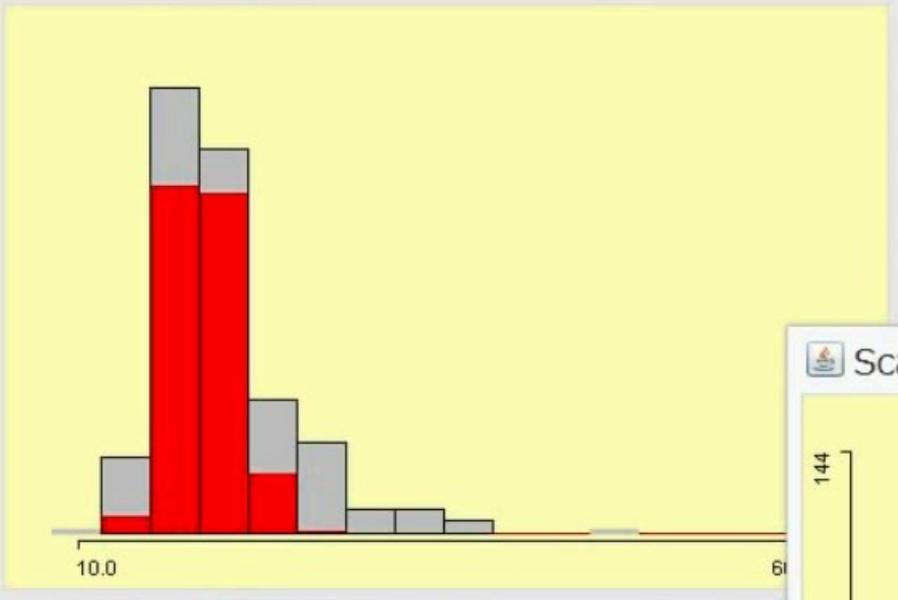


Mondrian(...)

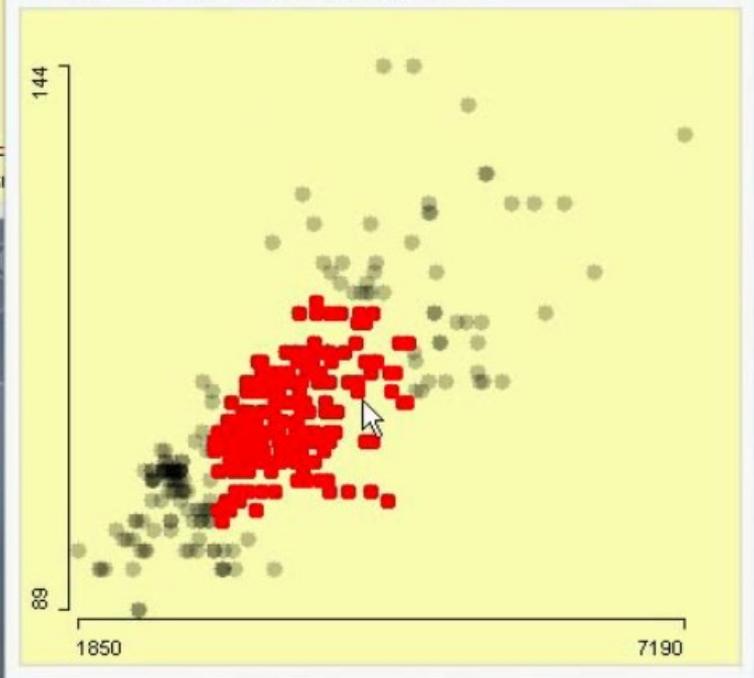
File Plot Calc Options Window Help

abc Vehic
abc Type
abc Drive
Sugg
Deale
Engin
Numb
Horse
City M
Highv
Weigl
Whee
Lengt
Width

Histogram(City Miles Per Gallon)



Scatterplot(x: Weight ...)



14:23

平行座標プロットとは

- 描き方
 - 連続変数の数だけの座標軸(縦方向)を横に平行に配置する
 - 座標軸の上端は最大値、下端は最小値とし、データの値をプロットする
 - 一つの個体の各座標軸上の点を折れ線で結ぶ
- 意味
 - データとして多いのは行が個体で列が変数の表
 - 平行座標プロットはそのデータの表をそのまま図にした
 - カテゴリー値も(ある程度)表せる
 - 外れ値を認識しやすい

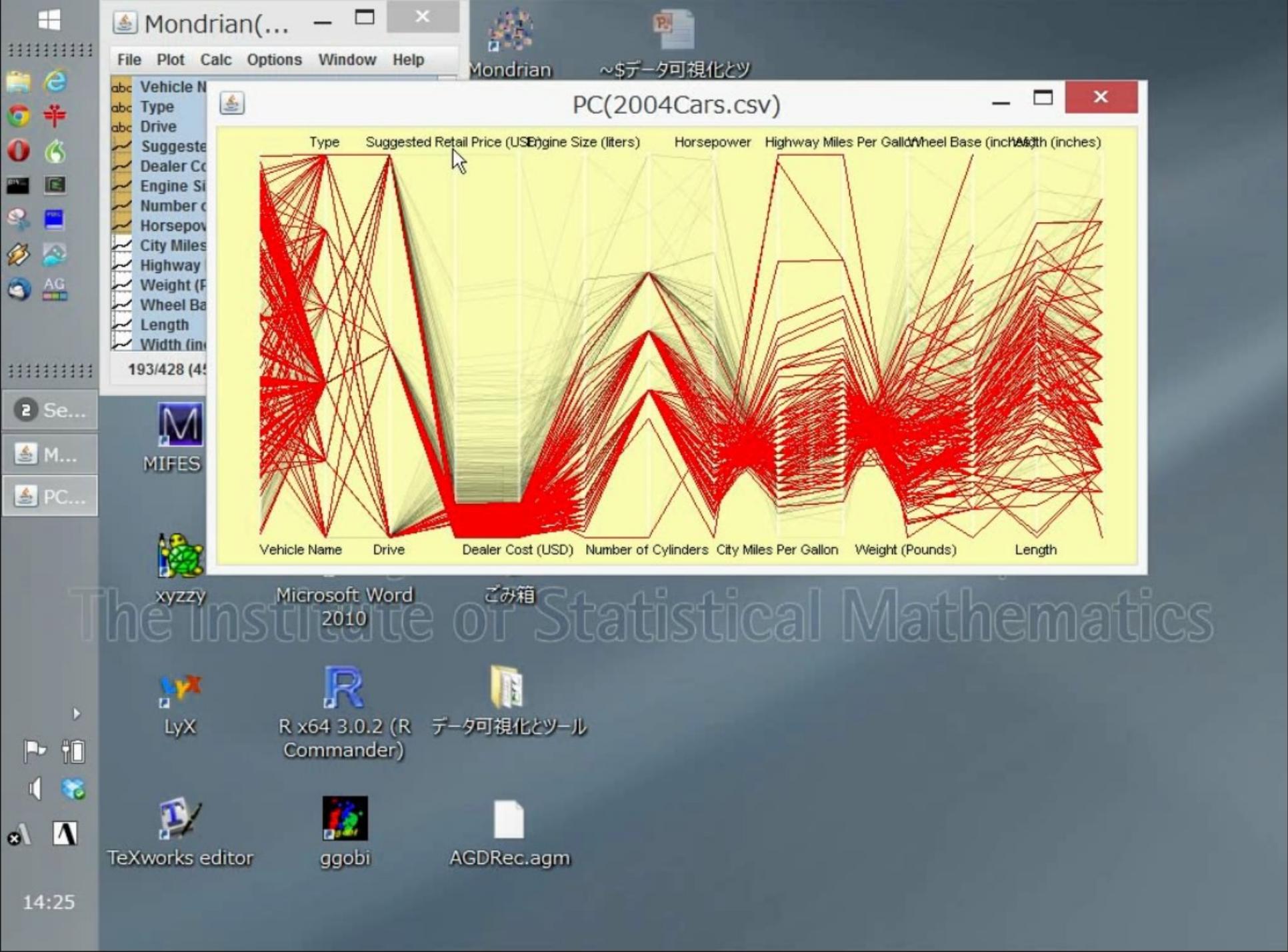


Mondrian(...)

File Plot Calc Options Window Help

- abc V Missing Value Plot
- abc T Barchart
- abc D Weighted Barchart
- D Histogram
- E Weighted Histogram
- N Scatterplot
- H Mosaic Plot
- V Weighted Mosaic Plot
- V **Parallel Coordinates**
- V Parallel Boxplot
- 270 Boxplot y by x
- SPLOM
- Map

search organization of Information and Systems
The Institute of Statistical Mathematics



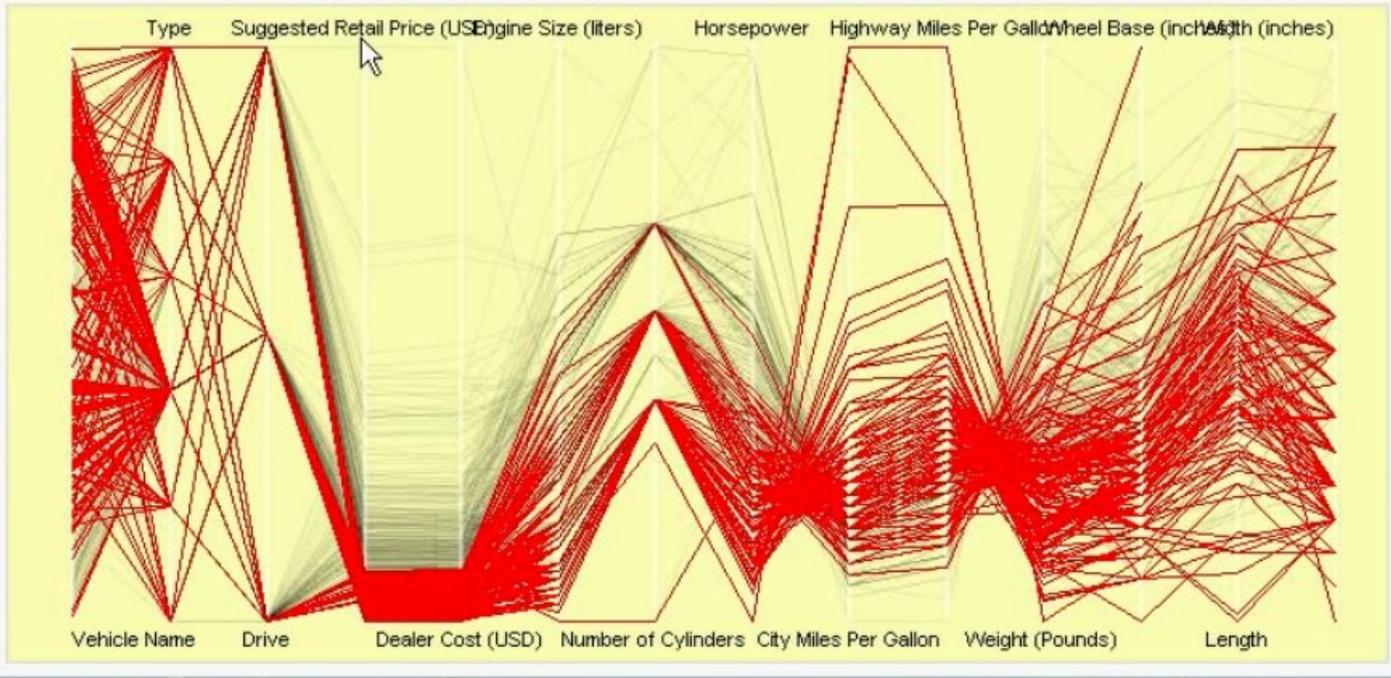
Mondrian(...)

File Plot Calc Options Window Help

Mondrian ~\$データ可視化とツ

PC(2004Cars.csv)

- abc Vehicle N
- abc Type
- abc Drive
- ~ Suggeste
- ~ Dealer Co
- ~ Engine Si
- ~ Number c
- ~ Horsepow
- ~ City Miles
- ~ Highway
- ~ Weight (P
- ~ Wheel Ba
- ~ Length
- ~ Width (in



- 2 Se...
- M...
- PC...



MIFES



xyzyzy

Microsoft Word 2010

ごみ箱



LyX



R x64 3.0.2 (R Commander)



データ可視化とツール



TeXworks editor



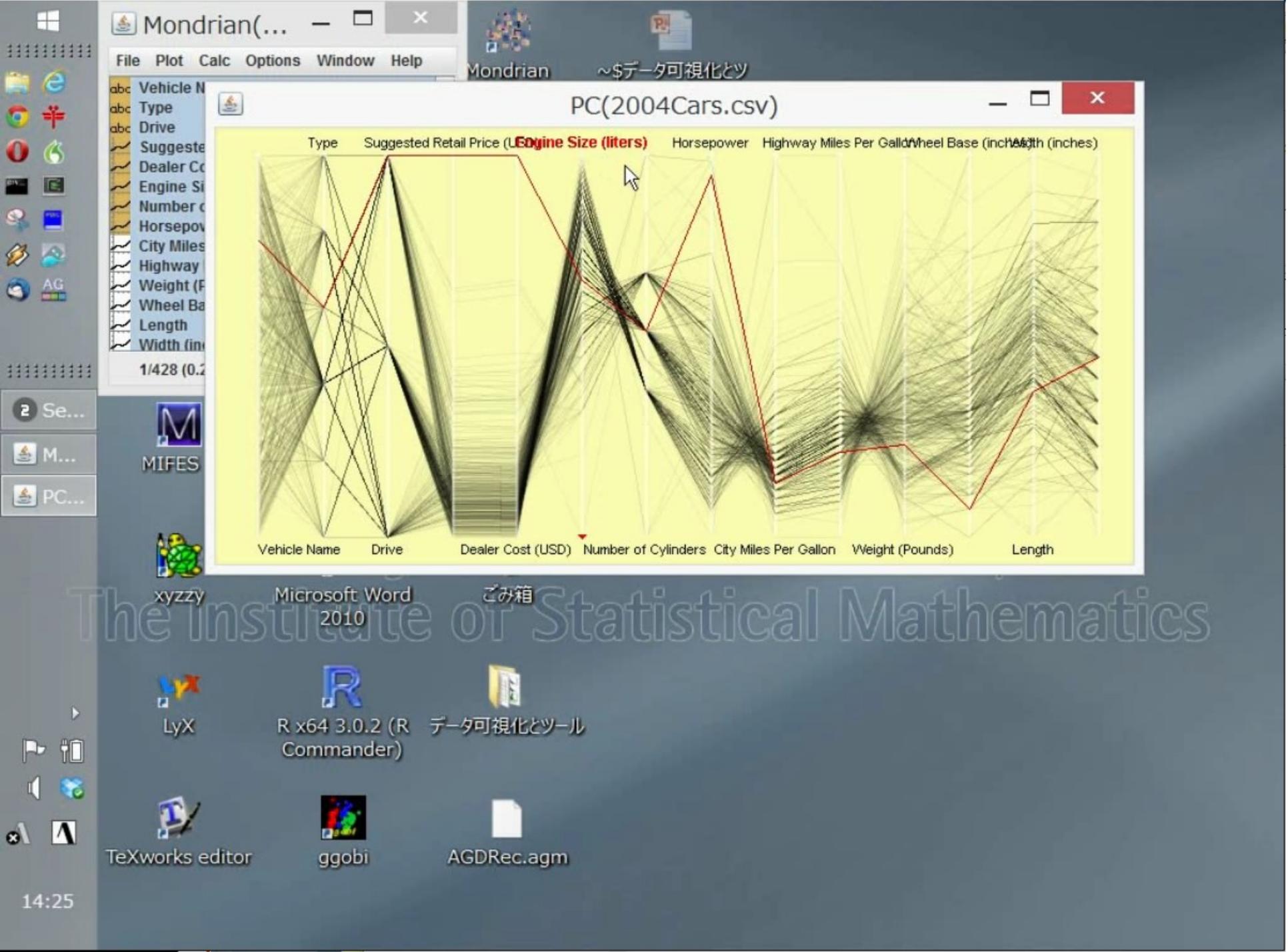
ggobi



AGDRec.agm

14:25

The Institute of Statistical Mathematics



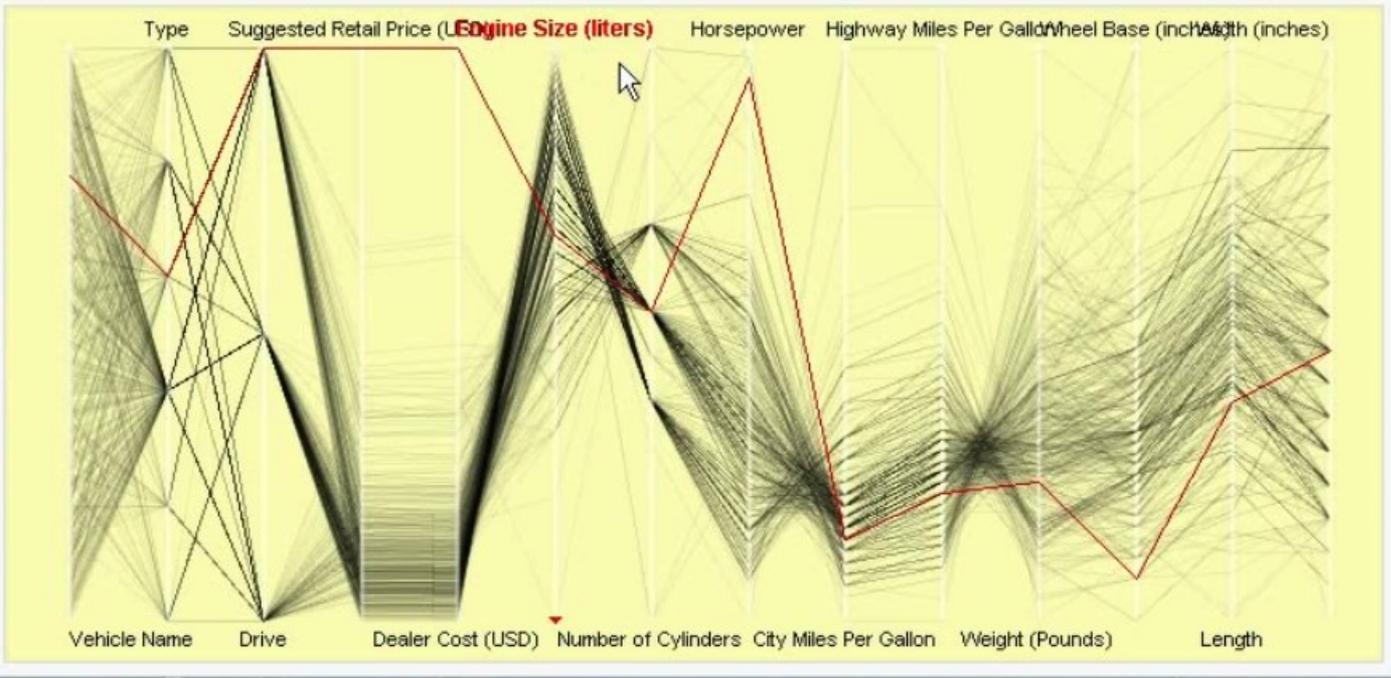
Mondrian(...)

File Plot Calc Options Window Help

Mondrian ~\$データ可視化とツ

PC(2004Cars.csv)

- abc Vehicle N
 - abc Type
 - abc Drive
 - Suggeste
 - Dealer Co
 - Engine Si
 - Number o
 - Horsepow
 - City Miles
 - Highway
 - Weight (P
 - Wheel Ba
 - Length
 - Width (in
- 1/428 (0.2



xyzyzy

Microsoft Word 2010

ごみ箱

LyX

R x64 3.0.2 (R Commander)

データ可視化とツール

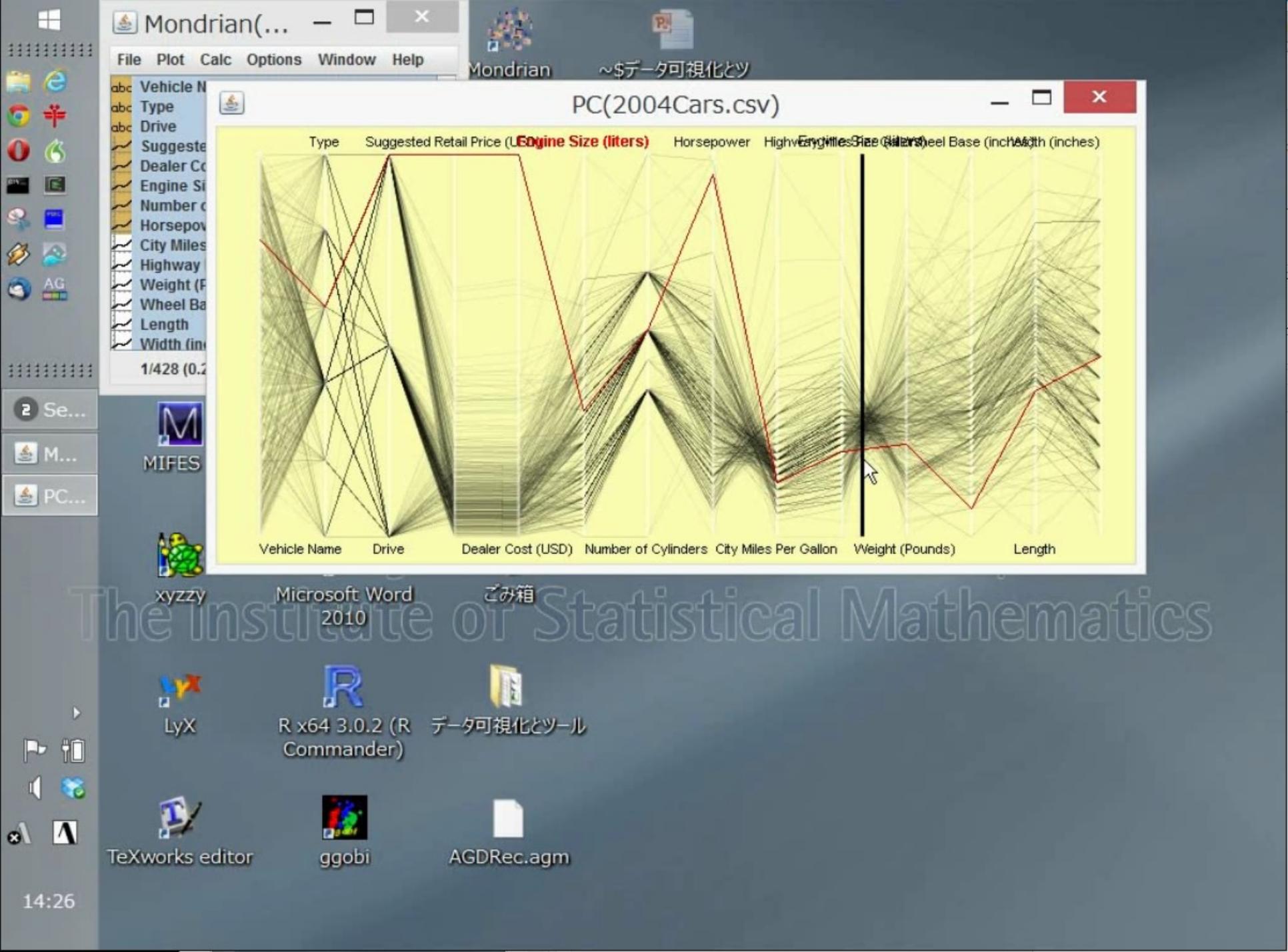
TeXworks editor

ggobi

AGDRec.agm

14:25

The Institute of Statistical Mathematics



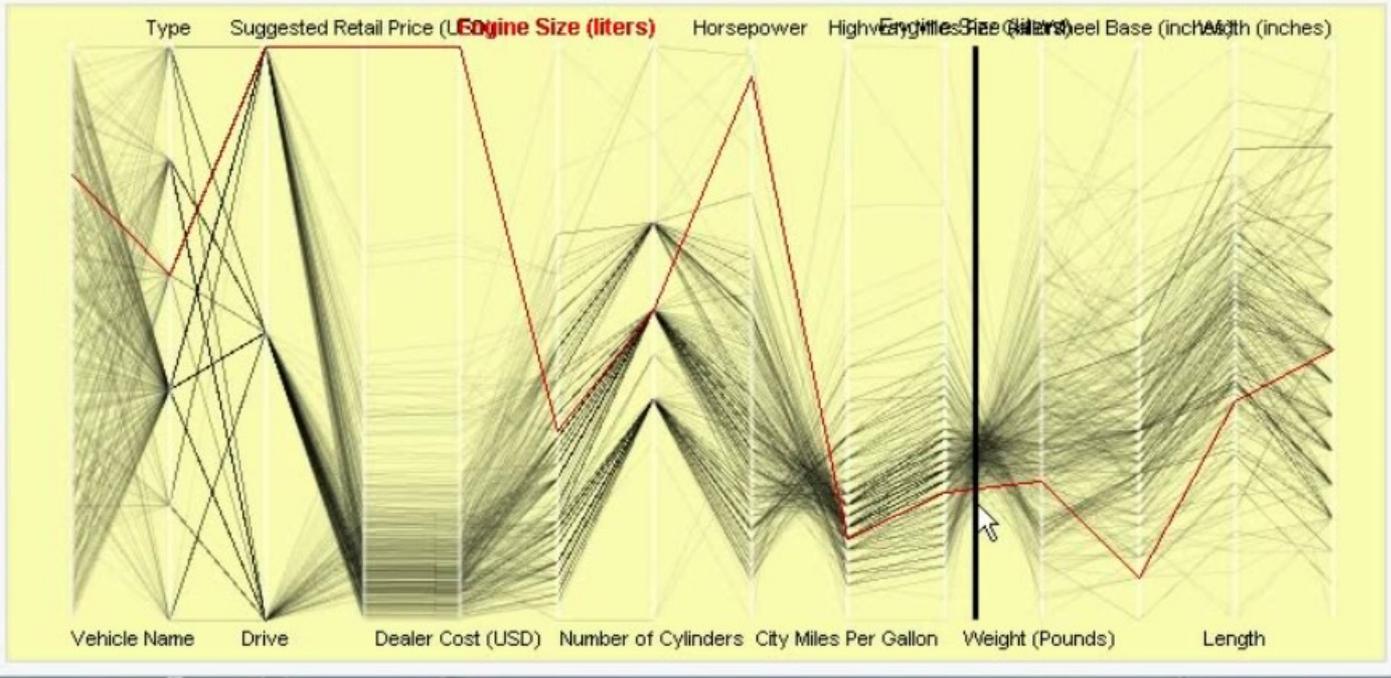
Mondrian(...)

File Plot Calc Options Window Help

Mondrian ~\$データ可視化とツ

PC(2004Cars.csv)

- abc Vehicle N
- abc Type
- abc Drive
- Suggeste
- Dealer Co
- Engine Si
- Number c
- Horsepow
- City Miles
- Highway
- Weight (P
- Wheel Ba
- Length
- Width (in



- 2 Se...
- M...
- PC...



MIFES



xyzyy

Microsoft Word 2010

ごみ箱



LyX



R x64 3.0.2 (R Commander)



データ可視化とツール



TeXworks editor



ggobi



AGDRec.agm

14:26

The Institute of Statistical Mathematics

平行座標プロットの対話的機能

- データ数が多くなると個体の区別が困難になる
 - 注目する個体を
選択し、強調表示する
ことにより区別できる
- 見やすくするための機能
 - 座標軸の配置の順序変更
 - 隣接する座標軸の変数の相関関係は見やすい
 - 座標軸の方向の変更
 - 軸方向を変えることで相関係数を逆転できる

モザイクプロットとは

- 描き方
 - 全体のデータを長方形で表す
 - 一つのカテゴリー変数の値でそれを縦方向に分割する
 - 次のカテゴリー変数の値で横方向に分割する
 - さらに次のカテゴリー変数の値で縦方向に分割
 - 以下、いくつかのカテゴリー変数を使って、横、縦、横、縦、、、のように交互に分割を続ける
- 意味
 - 複数のカテゴリー変数を持つデータの集計
(多重)分割表 = モザイクプロット

Mondrian(...)

File Plot Calc Options Window Help

- abc Class
- abc Age
- abc Sex
- abc Survived

0/2201 (0.00%)

Mondrian

~\$データ可視化とツ
ール.pptx

04Cars.csv

Titanic.txt

データ可視化とツ
ール.pptx

2 Se...

M...

M...

MIFES 9

Microsoft Excel 2010

データ可視化とツ
ール.pptx

xyzyzy

Microsoft Word 2010

ごみ箱

LyX

R x64 3.0.2 (R Commander)

データ可視化とツ
ール

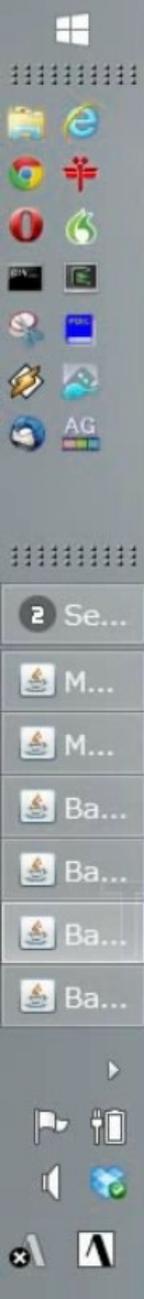
TeXworks editor

ggobi

AGDRec.agm



search organization of Information and Systems
The Institute of Statistical Mathematics



Mondrian(...)

File Plot Calc Options Window Help

abc Class
abc Age
abc Sex
abc Survived

0/2201 (0.00%)

Barchart(S...)

No

Yes

Barchart(Cl...)

Crew

First

Second

Third

Barchart(A...)

Adult

Child

Barchart(S...)

Female

Male

MIFES 9

Microsoft Excel 2010

データ可視化とツール.pptx

xyzzzy

Microsoft Word 2010

ごみ箱

LyX

R x64 3.0.2 (R Commander)

データ可視化とツール

TeXworks editor

ggobi

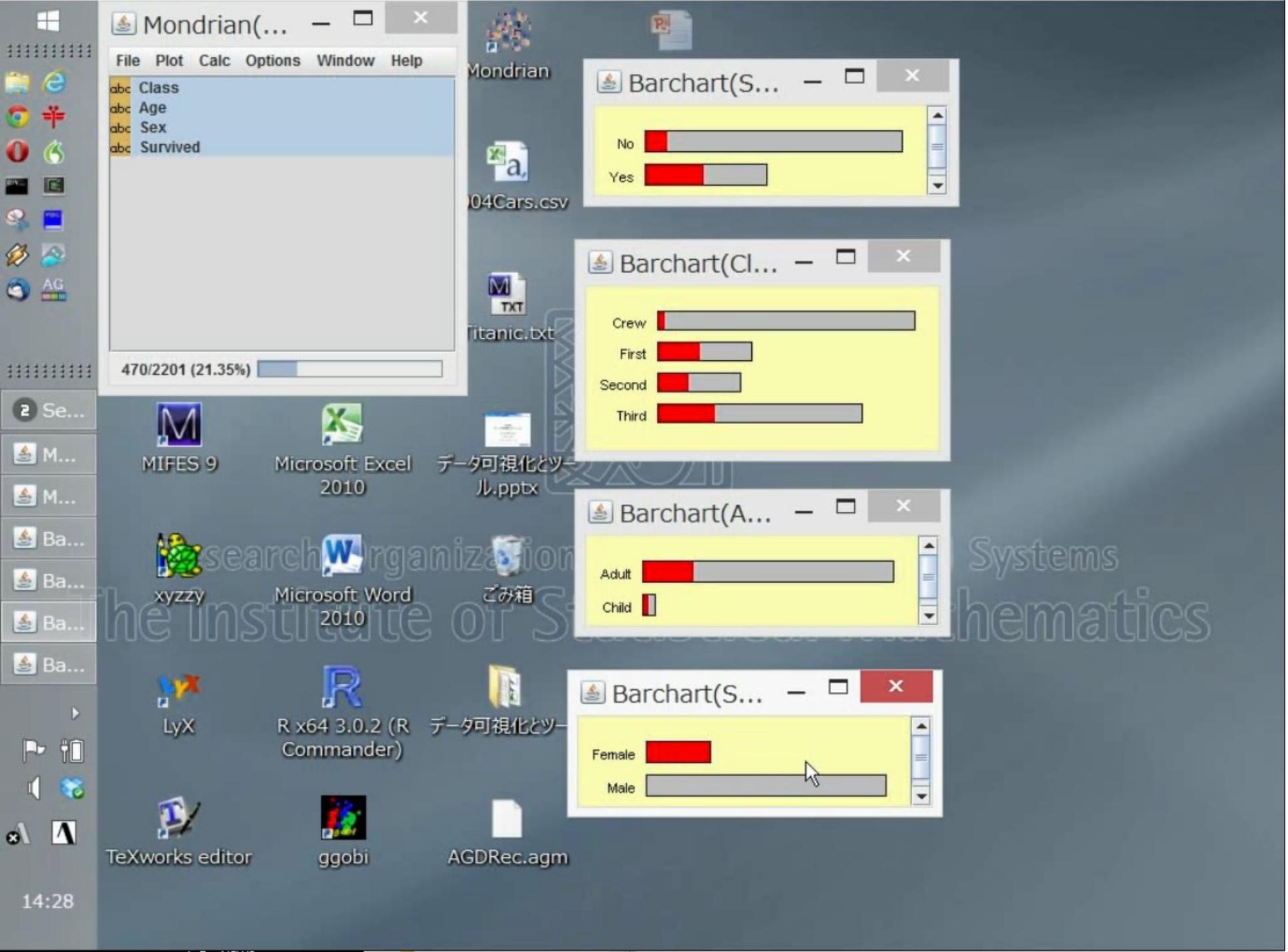
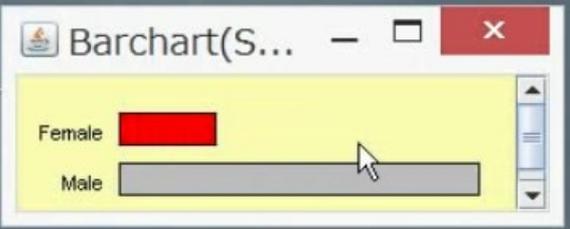
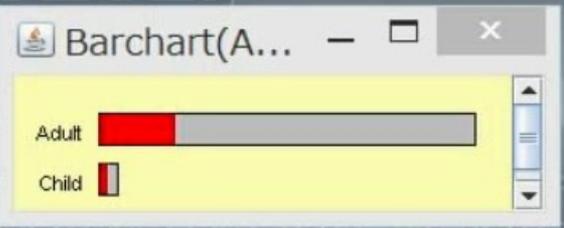
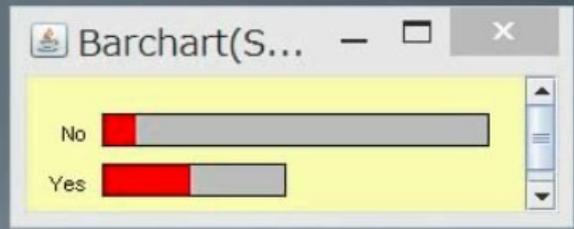
AGDRec.agm

Mondrian(...)

File Plot Calc Options Window Help

abc Class
abc Age
abc Sex
abc Survived

470/2201 (21.35%)



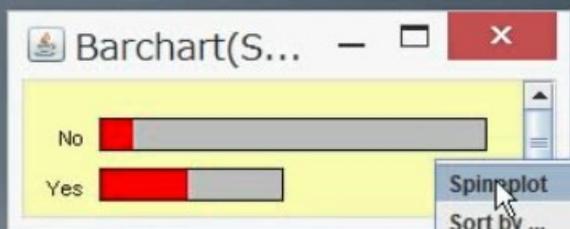
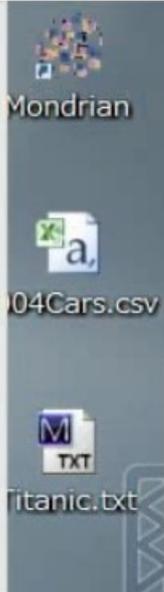


Mondrian(...)

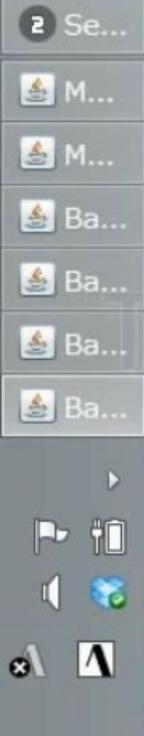
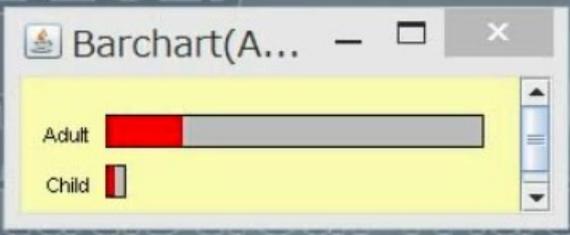
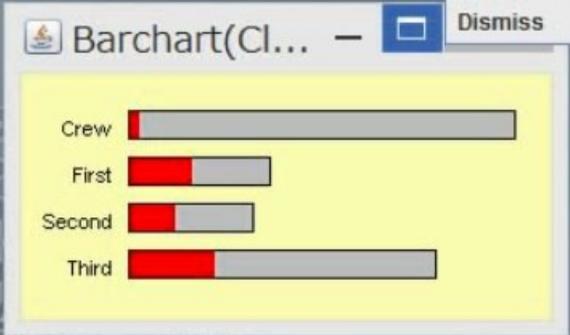
File Plot Calc Options Window Help

abc Class
abc Age
abc Sex
abc Survived

470/2201 (21.35%)



- Spin plot Ctrl-R
- Sort by ...
- Color Brush Ctrl-B
- Dismiss



MIFES 9

Microsoft Excel 2010

データ可視化とツール.pptx

xyzzzy

Microsoft Word 2010

ごみ箱

LyX

R x64 3.0.2 (R Commander)

データ可視化とツール

TeXworks editor

ggobi

AGDRec.agm

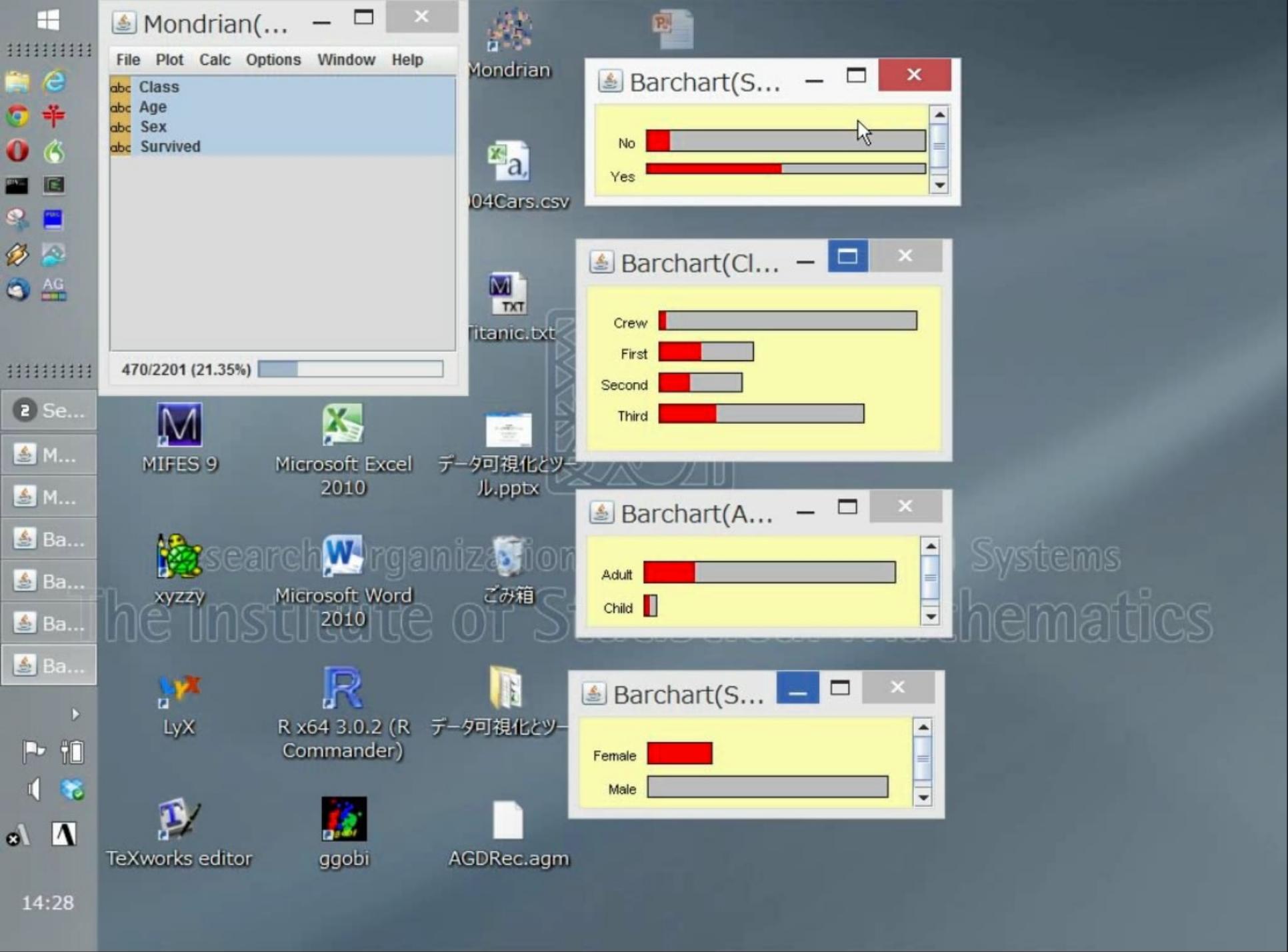
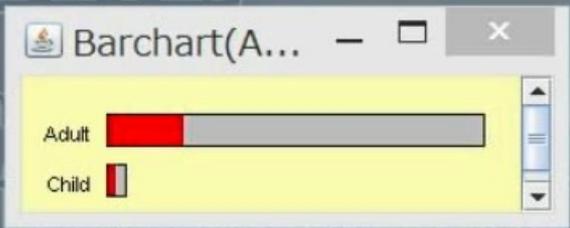
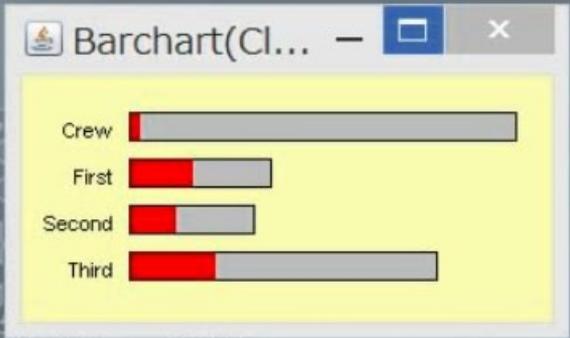
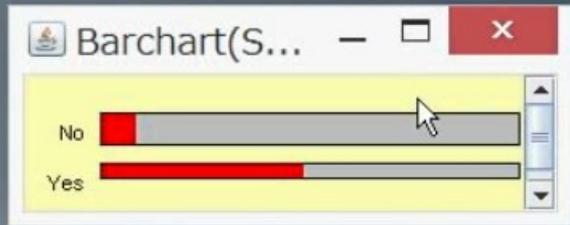
Systems
hematics

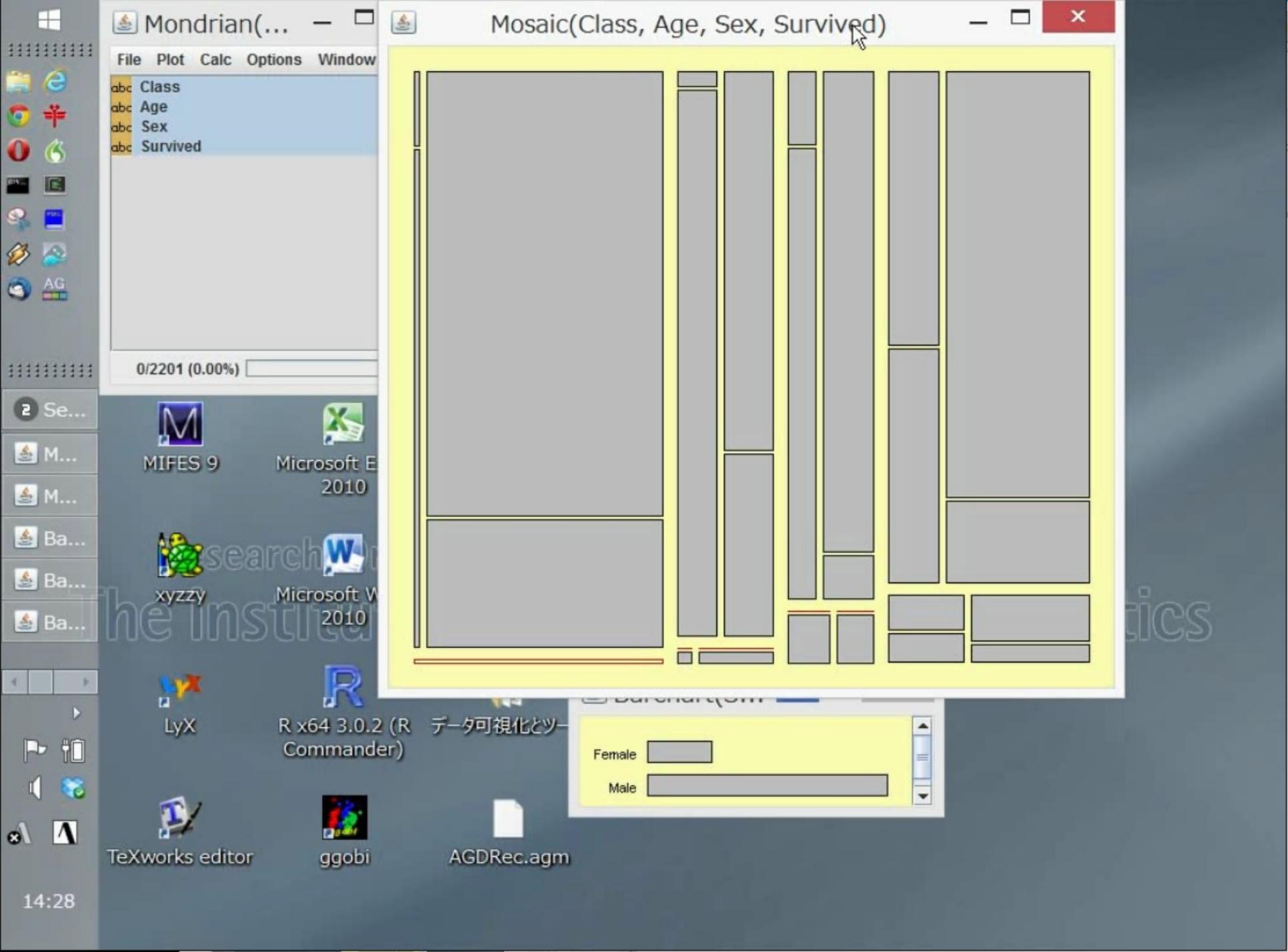
Mondrian(...)

File Plot Calc Options Window Help

abc Class
abc Age
abc Sex
abc Survived

470/2201 (21.35%)



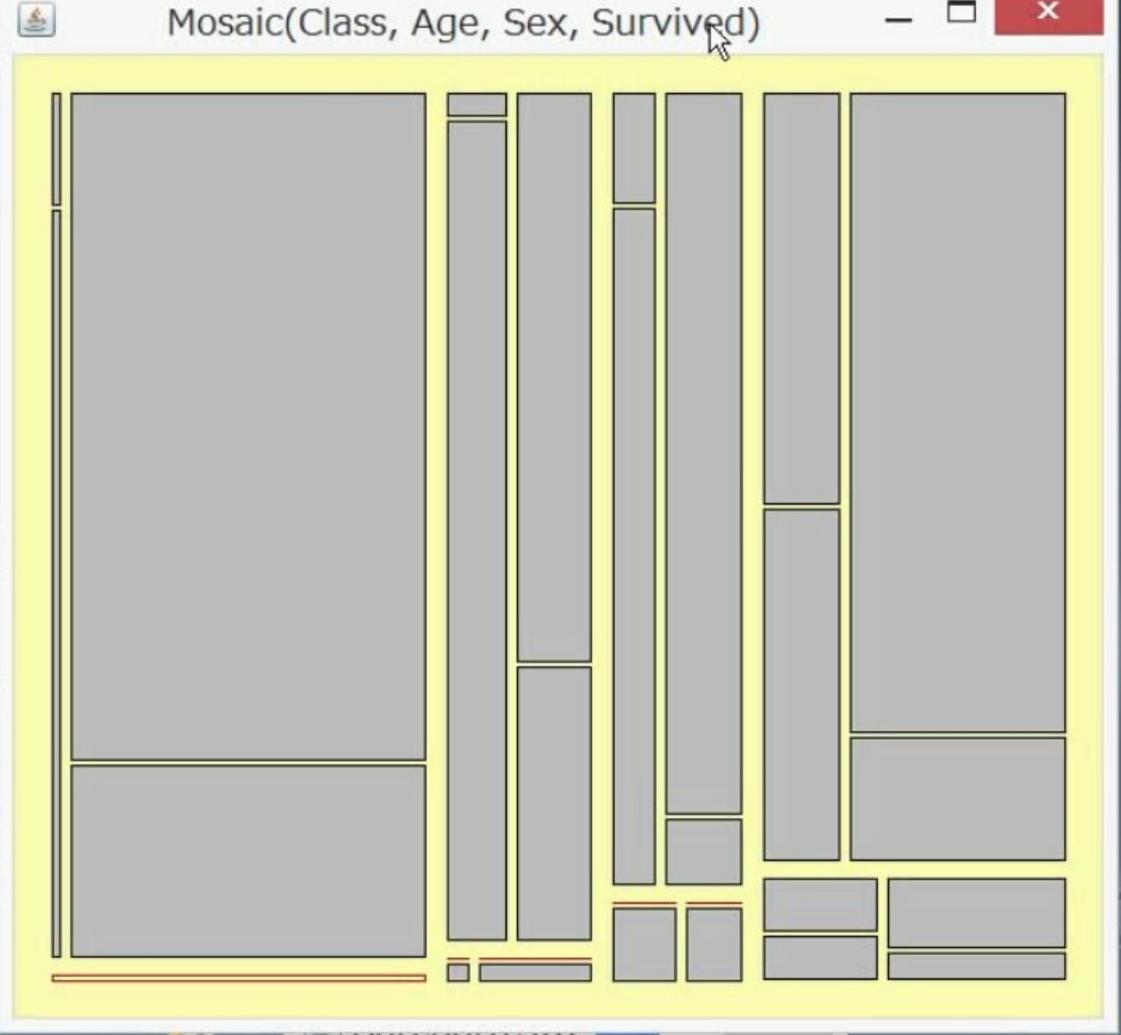


Mondrian(...)

File Plot Calc Options Window

- abc Class
- abc Age
- abc Sex
- abc Survived

0/2201 (0.00%)



Female

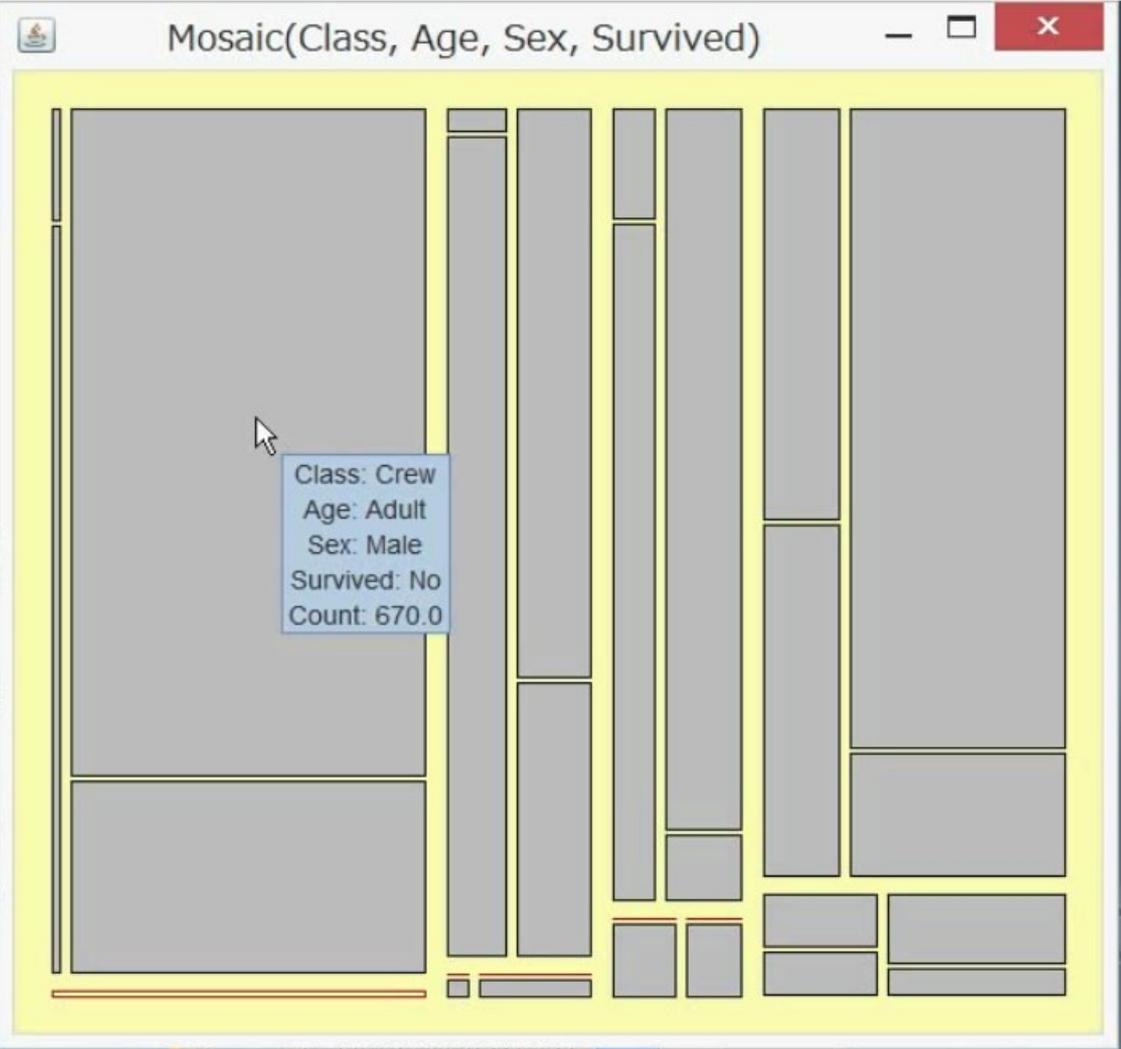
Male

Mondrian(...)

File Plot Calc Options Window

abc Class
abc Age
abc Sex
abc Survived

0/2201 (0.00%)



Se...

M...

M...

Ba...

Ba...

Ba...

MIFES 9

Microsoft E 2010

xyzzzy

Microsoft W 2010

LyX

R x64 3.0.2 (R Commander)

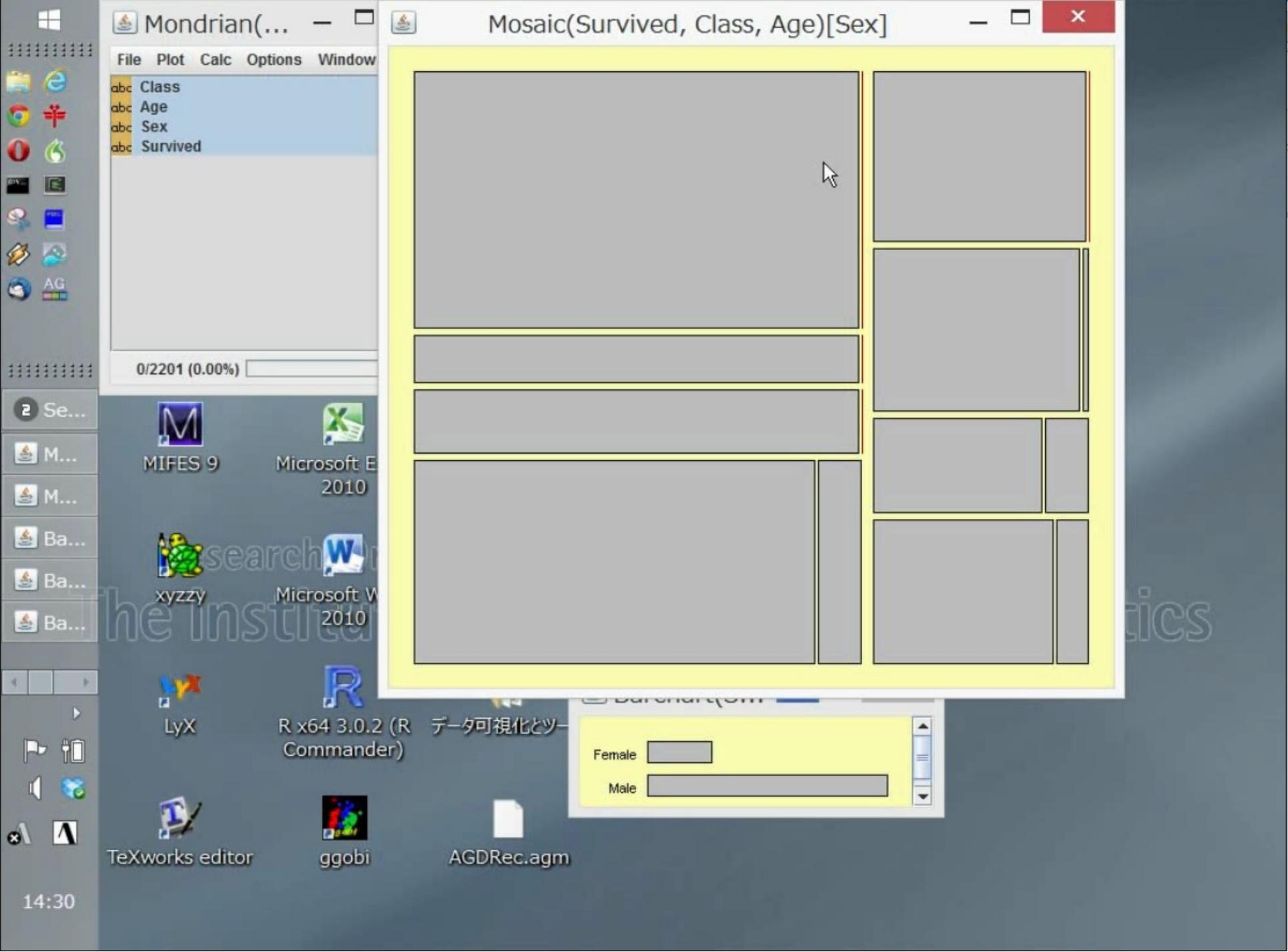
TeXworks editor

ggobi

AGDRec.agm

Female

Male



Mondrian(...)

Mosaic(Survived, Class, Age)[Sex]

File Plot Calc Options Window

- abc Class
- abc Age
- abc Sex
- abc Survived

0/2201 (0.00%)



MIFES 9



Microsoft E
2010



xyzyy



Microsoft W
2010



LyX



R x64 3.0.2 (R
Commander)



TeXworks editor



ggobi



AGDRec.agm

14:30

モザイクプロットの対話的機能

- モザイクプロットで重要なのは
 - 分割するための**カテゴリー変数の順序**
 - 分割するための**カテゴリー変数の個数**
- これらが容易に対話的に変更できると、データの重要な性質を発見しやすい
 - Mondrian では矢印キーで簡単に変更できる

参考文献

1. “The R Project for Statistical Computing”
<http://www.r-project.org/>
2. 山本義郎・飯塚誠也・藤野友和『Rで学ぶデータサイエンス
12. 統計データの視覚化』（共立出版、2013年）
3. “R Graphical Manual”
<http://rgm3.lab.nig.ac.jp/RGM/>
4. “GGobi data visualization system”
<http://www.ggobi.org/>
5. “Mondrian - Interactive Statistical Data Visualization in JAVA”
<http://www.rosuda.org/mondrian/>