EPSY 887: Computation Statistics

Plotting

Jason Bryer

http://github.com/jbryer/CompStats
jason@bryer.org

Week 4
February 18, 2013

Agenda

1. ggplot2: A Grammar of Graphics
2. likert Package

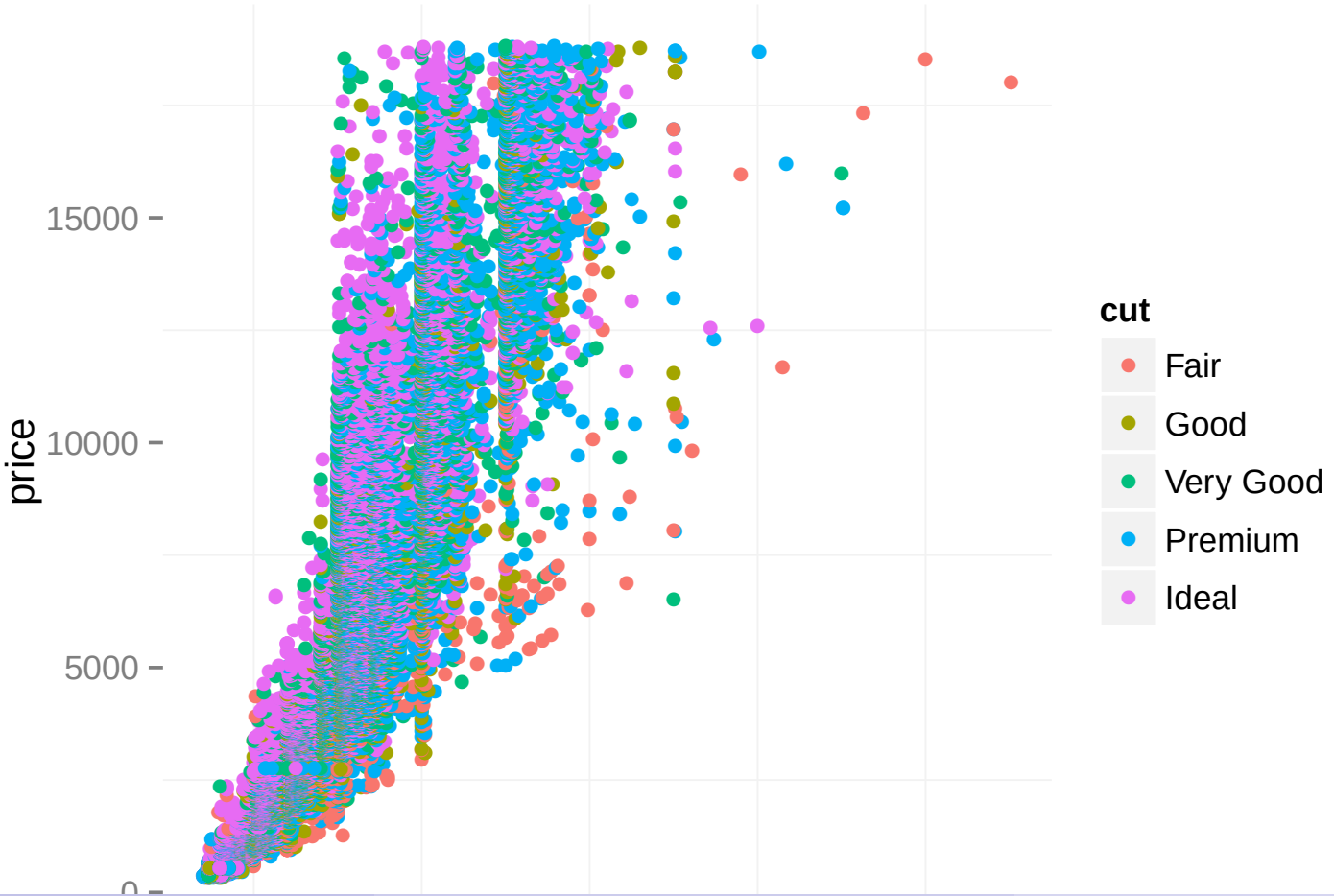
ggplot2: A Grammar of Graphics

- ggplot2 is an R package that provides an alternative framework based upon Wilkinson's (2005) Grammar of Graphics.
- ggplot2 is, in general, more flexible for creating "prettier" and complex plots.
- Works by creating layers of different types of objects/geometries (i.e. bars, points, lines, polygons, etc.)
- ggplot2 has at least three ways of creating plots:
  1. `qplot`
  2. `ggplot(...) + geom_XXX(...) + ...`
  3. `ggplot(...) + layer(...)`

  We will focus only on the second.

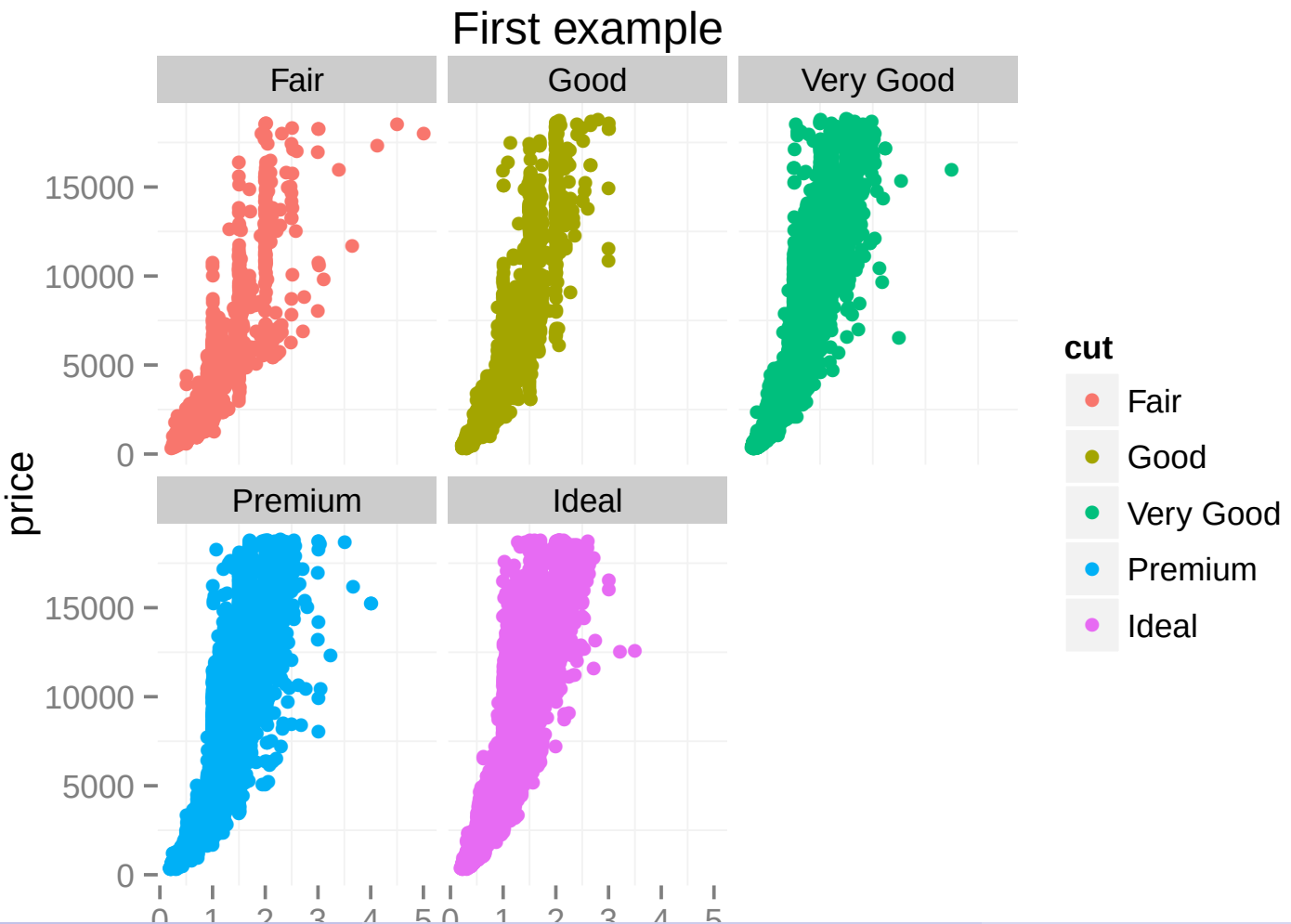
First Example

```
> data(diamonds)
> p <- ggplot(diamonds, aes(x=carat,y=price,colour=cut)) +
    geom_point()
> print(p)
```



First Example

```
> p <- p + facet_wrap(~cut) +
    ggtitle("First example")
> print(p)
```



Parts of a ggplot2 statement

- Data

```
ggplot(myDataFrame, aes(x=x, y=y)
```

Parts of a ggplot2 statement

- Data

  ```
  ggplot(myDataFrame, aes(x=x, y=y)
  ```

- Layers

  ```
  geom_point(), geom_histogram()
  ```

Parts of a ggplot2 statement

- Data

  `ggplot(myDataFrame, aes(x=x, y=y)`

- Layers

  `geom_point(), geom_histogram()`

- Facets

  `facet_wrap(~ cut), facet_grid(~ cut)`

Parts of a ggplot2 statement

- Data

  `ggplot(myDataFrame, aes(x=x, y=y)`

- Layers

  `geom_point(), geom_histogram()`

- Facets

  `facet_wrap(~ cut), facet_grid(~ cut)`

- Scales

  `scale_y_log10()`

Parts of a ggplot2 statement

- Data

  ```
  ggplot(myDataFrame, aes(x=x, y=y)
  ```

- Layers

  ```
  geom_point(), geom_histogram()
  ```

- Facets

  ```
  facet_wrap(~ cut), facet_grid(~ cut)
  ```

- Scales

  ```
  scale_y_log10()
  ```

- Other options

  ```
  ggtitle("my title"), ylim(c(0, 10000)), xlab("x-axis label")
  ```

Lots of geoms

```
geom_abline
geom_jitter
geom_area
geom_line
geom_bar
geom_linerange
geom_bin2d
geom_path
geom_blank
geom_point
geom_boxplot
geom_pointrange
geom_contour
geom_polygon
geom_crossbar
geom_quantile
```

```
geom_density
geom_rect
geom_density2d
geom_ribbon
geom_errorbar
geom_rug
geom_errorbarh
geom_segment
geom_freqpoly
geom_smooth
geom_hex
geom_step
geom_histogram
geom_text
geom_hline
geom_tile
geom_vline
```

Agenda

1. ggplot2: A Grammar of Graphics
2. likert Package

likert Package

The `likert` package provides functions to analyze and visualize Likert items. The graphics are created using `ggplot2`.

```
> require(devtools)
> install_github("likert","jbryer")
```

The package includes a subset of the PISA data. Item 28 of the student questionnaire contains 11 items about student reading habits.

```
> require(likert)
> data(pisaitems)
> items28 = pisaitems[,substr(names(pisaitems), 1,5) == "ST24Q"]
```

Data Setup

First, rename the columns to the item stems.

```
> items28 <- rename(items28, c(
    ST24Q01="I read only if I have to.",
    ST24Q02="Reading is one of my favorite hobbies.",
    ST24Q03="I like talking about books with other people.",
    ST24Q04="I find it hard to finish books.",
    ST24Q05="I feel happy if I receive a book as a present.",
    ST24Q06="For me, reading is a waste of time.",
    ST24Q07="I enjoy going to a bookstore or a library.",
    ST24Q08="I read only to get information that I need.",
    ST24Q09="I cannot sit still and read for more than a few minutes
    ST24Q10="I like to express my opinions about books I have read."
    ST24Q11="I like to exchange books with my friends"))
```

Analyzing Likert Items

The `likert` function will analyze the items.

```
> l28 = likert(items28)
```

And the `print` method will provide percentages of each category.

```
> print(l28)
```

```
                                                    Item
1                                I read only if I have to.
2                     Reading is one of my favorite hobbies.
3              I like talking about books with other people.
4                            I find it hard to finish books.
5             I feel happy if I receive a book as a present.
6                       For me, reading is a waste of time.
7                I enjoy going to a bookstore or a library.
8                I read only to get information that I need.
9  I cannot sit still and read for more than a few minutes.
10   I like to express my opinions about books I have read.
11                  I like to exchange books with my friends
   Strongly disagree Disagree Agree Strongly agree
1                 23       36    31           10.7
2                 20       36    32           11.4
3                 21       34    36            9.0
4                 25       40    27            8.1
5                 19       28    40           12.9
6                 42       41    11            6.1
7                 18       33    37           11.9
8                 15       35    36           13.8
9                 33       43    17            6.8
10                14       28    44           15.1
11                23       33    32           12.4
```

Likert SUmmary

The summary method will provide provide descriptive statistics.

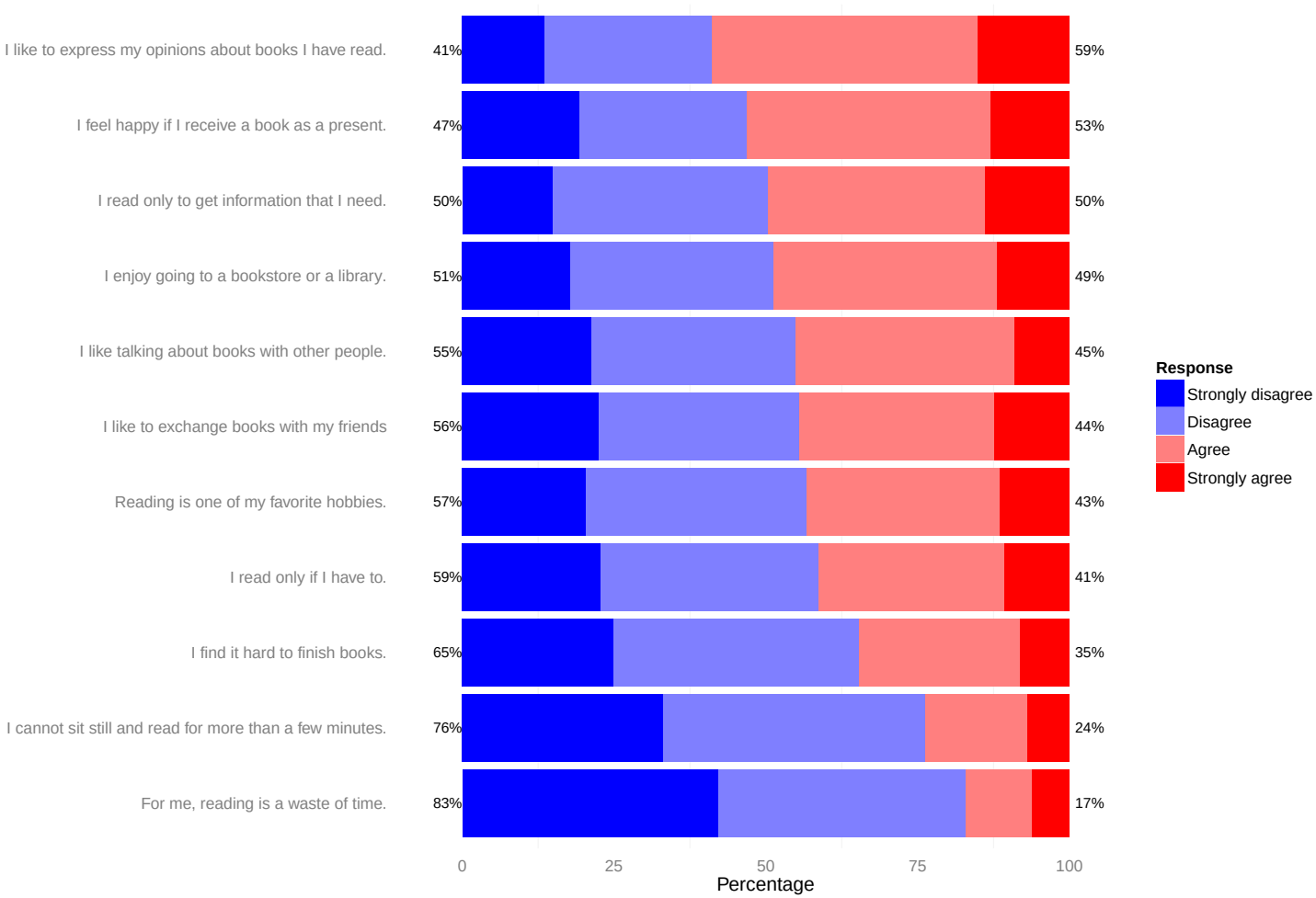
```
> summary(128)
```

	Item	low	high	mean	sd
1	I read only if I have to.	59	41	2.3	0.94
2	Reading is one of my favorite hobbies.	57	43	2.3	0.93
3	I like talking about books with other people.	55	45	2.3	0.91
4	I find it hard to finish books.	65	35	2.2	0.90
5	I feel happy if I receive a book as a present.	47	53	2.5	0.94
6	For me, reading is a waste of time.	83	17	1.8	0.86
7	I enjoy going to a bookstore or a library.	51	49	2.4	0.92
8	I read only to get information that I need.	50	50	2.5	0.91
9	I cannot sit still and read for more than a few minutes.	76	24	2.0	0.88
10	I like to express my opinions about books I have read.	41	59	2.6	0.90
11	I like to exchange books with my friends	56	44	2.3	0.96

Bar Plot

The `summary` method will provide provide descriptive statistics.

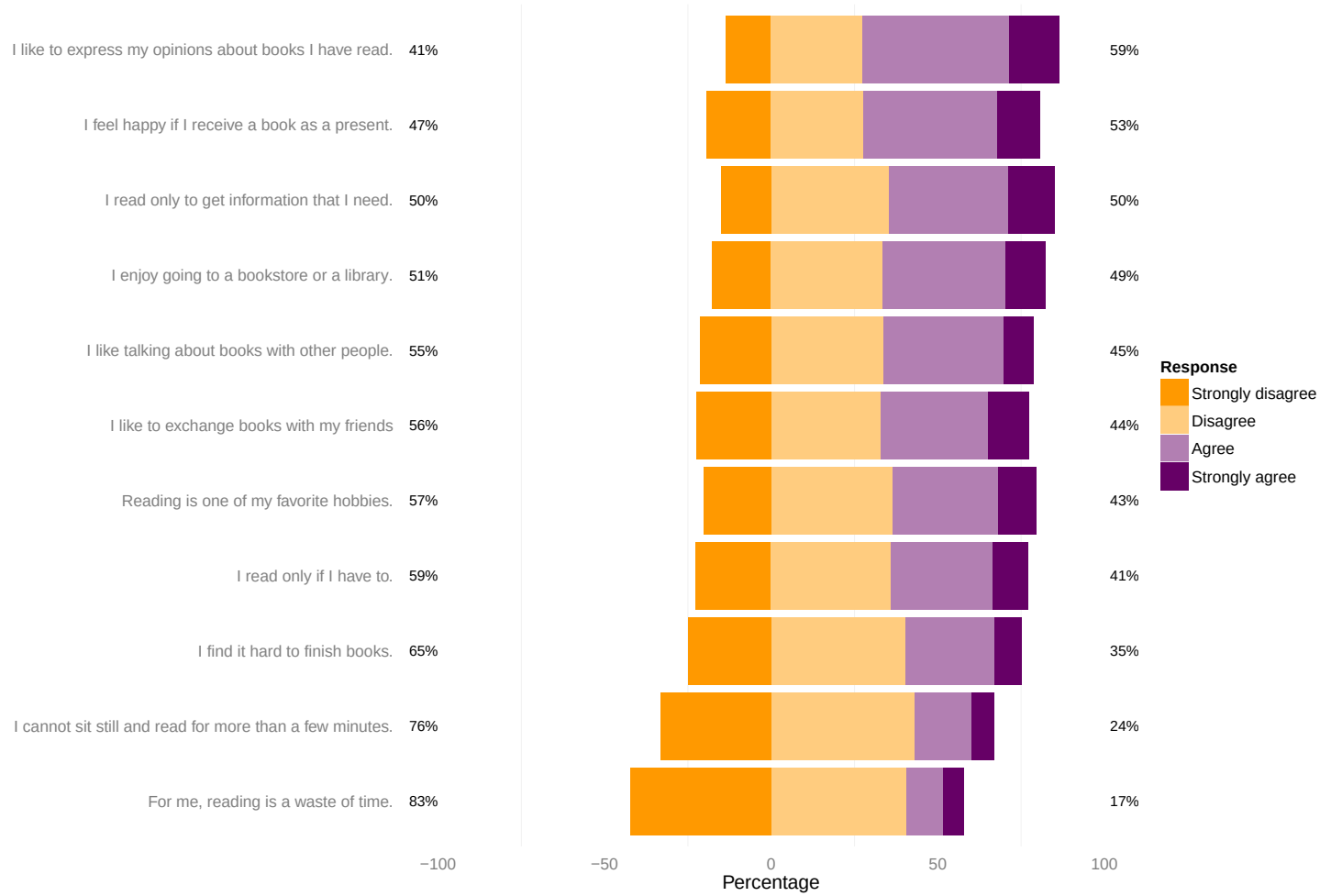
```
> print(plot(128))
```



Bar Plot Centered

The `summary` method will provide provide descriptive statistics.

```
> print(plot(l28, centered=TRUE, low.color="#FF9900", high.color="#6
    )
```



Heat Map

The `summary` method will provide provide descriptive statistics.

```
> print(plot(l28, type="heat")
    )
```