

Feature Tour

By Benedict Escoto



This paper is produced mechanically as part of FAViR.
See <http://www.favir.net> for more information.

Abstract

This paper intended as an initial tour of the functions included in the FAViR package. It itself is an example of a FAViR paper that uses these functions.

1 Introduction

The R package component of FAViR includes several formatting functions which make writing FAViR papers more convenient. They are covered in the package manual pages but some only make sense inside a FAViR paper. This paper demonstrates each of these functions and can be used as a quick-reference or a feature tour of the FAViR package.

When reading the PDF version of this paper, please follow along in the associated Sweave (.Rnw) version.

2 LaTeX

LaTeX commands are available inside the LaTeX (non-R) sections of a FAViR paper. In particular:

- *text in italics*
- **boldface text**
- `typewriter font`
- tiny script footnote small normal large Large LARGE huge Huge

This text is centered.

1. This list is numbered.

2. And it has equations like $3 \leq 4$ and $1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots = \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$

You can also have numbered, centered equations, like equation (1) below.

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2/5)/\sqrt{5} & |x| < \sqrt{5} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

LaTeX has far too many features to demonstrate here, so please consult the internet for general LaTeX help.

3 Sweave

Sweave allows R code to be included in a paper and automatically executed. Below is the most basic Sweave mode, where both R commands and their output are printed:

```
> 2 + 2
```

```
[1] 4
```

```
> 1:10
```

```
[1] 1 2 3 4 5 6 7 8 9 10
```

However, the output of this mode generally doesn't look very polished. A more powerful technique is not to display the R commands, but have them generate LaTeX output directly. This is how most code segments in FAViR papers are run.

This LaTeX is generated by R

$$\sum_{i=1}^{100} \log(i) \approx 363.739375555563$$

Note the double backslashes in the above R code. When that code is evaluated by R, each double backslash becomes a single backslash which is then evaluated by LaTeX.

It's also possible to run R without displaying the results or even in the *middle* of a sentence.

4 FAViR Data Frames

This section is about presenting tabular data in R data frames. First we'll define a simple data frame and then progressively refine its presentation by using FAViR functions.

An easy way to present a data frame is by using basic Sweave mode:

```
> df <- data.frame(year = 2000:2008, prem.a = 2e+05 * rnorm(9,
+   1, 0.1), loss.a = 130000 * rnorm(9, 1, 0.2), loss.ratio.a = 0,
+   prem.b = 50000 * rnorm(9, 1, 0.1), loss.b = 30000 * rnorm(9,
+   1, 0.2), loss.ratio.b = 0)
> df$loss.ratio.a <- df$loss.a/df$prem.a
> df$loss.ratio.b <- df$loss.b/df$prem.b
> df
```

	year	prem.a	loss.a	loss.ratio.a	prem.b	loss.b	loss.ratio.b
1	2000	183498.0	129739.18	0.7070332	55025.13	27297.34	0.4960886
2	2001	212155.6	107513.20	0.5067657	37027.59	26323.99	0.7109291
3	2002	196266.6	123500.99	0.6292510	50558.92	34621.84	0.6847822
4	2003	204301.0	140802.65	0.6891923	47320.00	25544.11	0.5398164
5	2004	160060.5	143219.47	0.8947833	42653.64	34947.62	0.8193349
6	2005	217680.0	159874.57	0.7344476	49885.65	28446.66	0.5702374
7	2006	208696.3	162417.95	0.7782503	55589.48	30327.70	0.5455655
8	2007	174659.7	145887.79	0.8352687	37513.22	31433.95	0.8379433
9	2008	209397.5	91271.52	0.4358768	50663.04	32016.73	0.6319544

However the results are obviously rough. We can improve on this by using a FavirDF object:

year	prem.a	loss.a	loss.ratio.a	prem.b	loss.b	loss.ratio.b
2,000	183,498	129,739	1	55,025	27,297	0
2,001	212,156	107,513	1	37,028	26,324	1
2,002	196,267	123,501	1	50,559	34,622	1
2,003	204,301	140,803	1	47,320	25,544	1
2,004	160,061	143,219	1	42,654	34,948	1
2,005	217,680	159,875	1	49,886	28,447	1
2,006	208,696	162,418	1	55,589	30,328	1
2,007	174,660	145,888	1	37,513	31,434	1
2,008	209,398	91,272	0	50,663	32,017	1

Figure 1:

The results are now displayed in a figure. LaTeX does not necessarily display figures where they are defined—it uses an algorithm to place them where it thinks they will fit. It's good practice to add labels and captions to each FavirDF, so it's clear which table is being referred to.

year	prem.a	loss.a	loss.ratio.a	prem.b	loss.b	loss.ratio.b
2,000	183,498	129,739	1	55,025	27,297	0
2,001	212,156	107,513	1	37,028	26,324	1
2,002	196,267	123,501	1	50,559	34,622	1
2,003	204,301	140,803	1	47,320	25,544	1
2,004	160,061	143,219	1	42,654	34,948	1
2,005	217,680	159,875	1	49,886	28,447	1
2,006	208,696	162,418	1	55,589	30,328	1
2,007	174,660	145,888	1	37,513	31,434	1
2,008	209,398	91,272	0	50,663	32,017	1

Figure 2: This FavirDF has a label and caption.

Figure 2 has a label and caption. It arguably looks better than the simple Sweave version, but the headers are chosen for R programming convenience and are not ideal for presentation.

4.1 Table Headings

We can change the headings by using the `FieldHeadings` function. Figure 3 has better column headings.

In figure 4 we do even better by grouping the lines of business.

Calendar Year	Earned Premium Line A	Ultimate Loss Line A	Loss Ratio Line A	Earned Premium Line B	Ultimate Loss Line B	Loss Ratio Line B
2,000	183,498	129,739	1	55,025	27,297	0
2,001	212,156	107,513	1	37,028	26,324	1
2,002	196,267	123,501	1	50,559	34,622	1
2,003	204,301	140,803	1	47,320	25,544	1
2,004	160,061	143,219	1	42,654	34,948	1
2,005	217,680	159,875	1	49,886	28,447	1
2,006	208,696	162,418	1	55,589	30,328	1
2,007	174,660	145,888	1	37,513	31,434	1
2,008	209,398	91,272	0	50,663	32,017	1

Figure 3: Table with better column headings

Calendar Year	Line of Business A			Line of Business B		
	Earned Premium	Ultimate Loss	Loss Ratio	Earned Premium	Ultimate Loss	Loss Ratio
2,000	183,498	129,739	1	55,025	27,297	0
2,001	212,156	107,513	1	37,028	26,324	1
2,002	196,267	123,501	1	50,559	34,622	1
2,003	204,301	140,803	1	47,320	25,544	1
2,004	160,061	143,219	1	42,654	34,948	1
2,005	217,680	159,875	1	49,886	28,447	1
2,006	208,696	162,418	1	55,589	30,328	1
2,007	174,660	145,888	1	37,513	31,434	1
2,008	209,398	91,272	0	50,663	32,017	1

Figure 4: Table with group headings

4.2 Entry Formatting

This table is starting to look decent, but the numbers are not formatted correctly. You can specify arbitrary formatting functions per field, but several standard ones are built into the FAViR package. In figure 5 the loss ratio and years columns have been reformatted.

Calendar Year	Line of Business A			Line of Business B		
	Earned Premium	Ultimate Loss	Loss Ratio (%)	Earned Premium	Ultimate Loss	Loss Ratio (%)
2000	183,498	129,739	70.7	55,025	27,297	49.6
2001	212,156	107,513	50.7	37,028	26,324	71.1
2002	196,267	123,501	62.9	50,559	34,622	68.5
2003	204,301	140,803	68.9	47,320	25,544	54.0
2004	160,061	143,219	89.5	42,654	34,948	81.9
2005	217,680	159,875	73.4	49,886	28,447	57.0
2006	208,696	162,418	77.8	55,589	30,328	54.6
2007	174,660	145,888	83.5	37,513	31,434	83.8
2008	209,398	91,272	43.6	50,663	32,017	63.2

Figure 5: Table with formatted entries

Note also that LaTeX moved some of the earlier figures onto other pages because there were “too many” figures and not enough text. At the start of this section we have manually forced all figures to be placed in order to start fresh.

4.3 Final Table

The last change we will make is to add a summary row. Figure 6 demonstrates this.

Calendar Year	Line of Business A			Line of Business B		
	Earned Premium	Ultimate Loss	Loss Ratio (%)	Earned Premium	Ultimate Loss	Loss Ratio (%)
2000	183,498	129,739	70.7	55,025	27,297	49.6
2001	212,156	107,513	50.7	37,028	26,324	71.1
2002	196,267	123,501	62.9	50,559	34,622	68.5
2003	204,301	140,803	68.9	47,320	25,544	54.0
2004	160,061	143,219	89.5	42,654	34,948	81.9
2005	217,680	159,875	73.4	49,886	28,447	57.0
2006	208,696	162,418	77.8	55,589	30,328	54.6
2007	174,660	145,888	83.5	37,513	31,434	83.8
2008	209,398	91,272	43.6	50,663	32,017	63.2
Total	1,766,715	1,204,227	68.2	426,237	270,960	63.6

Figure 6: Final table with summary row

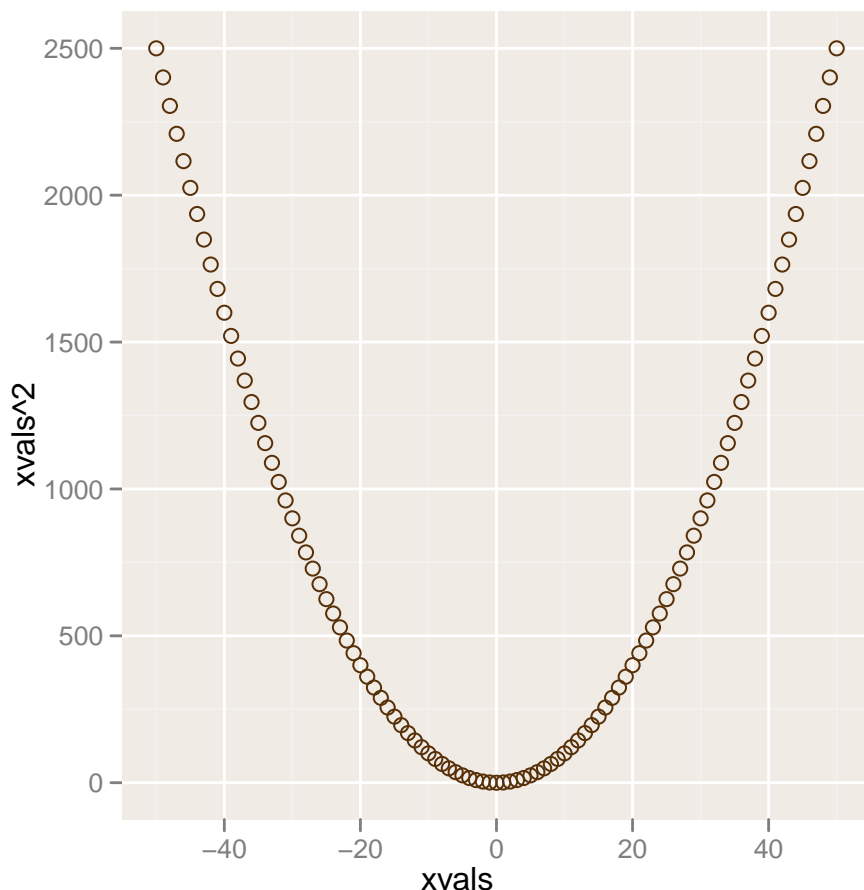


Figure 7: Basic graph with ggplot2

5 Graphics

ggplot2 is an advanced graphing package for R. The “gg” stands for “grammar of graphics”. The idea is that you just need to learn a few basic graphing concepts (such as geoms, aesthetics, and scales) and then you can combine them in myriad ways to create great custom plots. There is lots of documentation for *ggplot2*; this paper just demonstrates the FAViR functions which include them into a paper.

Figure 7 is a simple graph made with *ggplot2*.

The size can be adjusted with the **height** and **width** parameters as shown in figure 8.

Figure 9 shows how you can easily put multiple graphs into different viewports with the **IncludeGrid** function. However, if the graphs all have comparable values in them, it’s probably better to use *ggplot2*’s powerful faceting functionality.

Our final feature demonstration is to automatically include some legal boilerplate in the

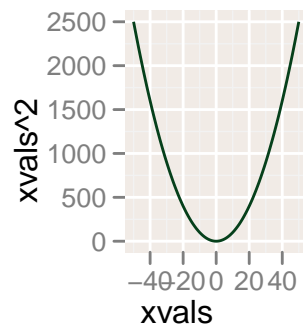


Figure 8: Smaller Graph

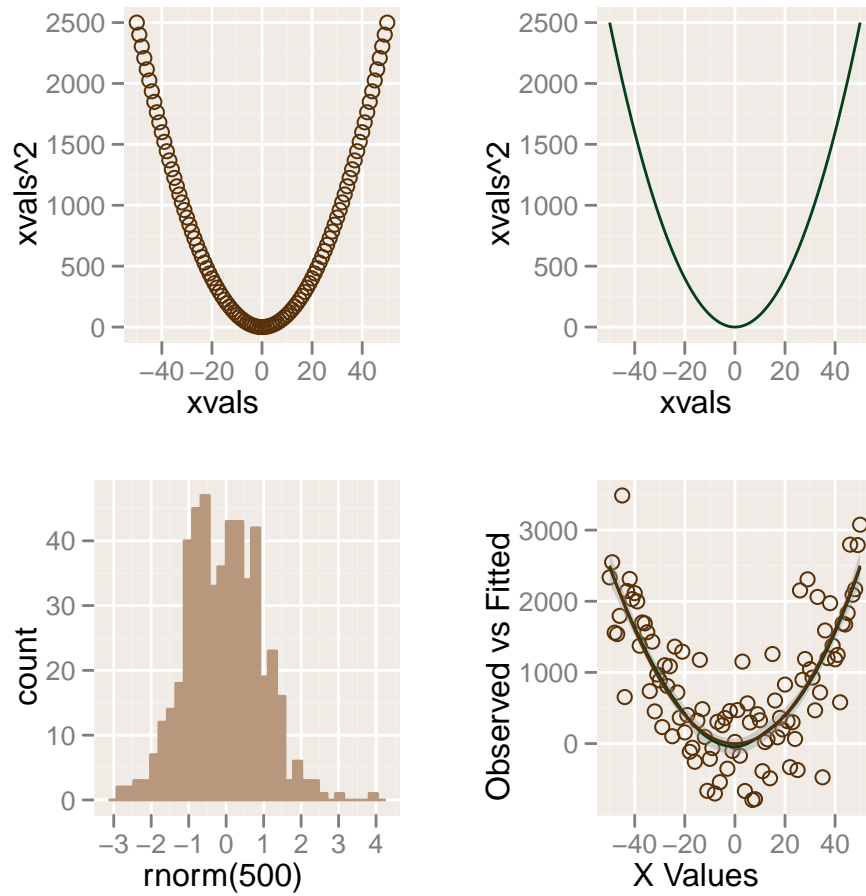


Figure 9: Four separate plots in one grid

next section.

6 Legal

Copyright © 2010 FAViR Project

This paper is part of the FAViR project. All the R source code used to produce it is freely distributable under the GNU General Public License. See <http://www.favir.net> for more information on FAViR or to download the source code for this paper.

Copying and distribution of this paper itself, with or without modification, are permitted in any medium without royalty provided the copyright notice and this notice are preserved. This paper is offered as-is, without any warranty.

This paper is intended for educational purposes only and should not be used to violate anti-trust law. The authors and FAViR editors do not necessarily endorse the information or techniques in this paper and assume no responsibility for their accuracy.