

We often encounter nightmarish scenario while generating final versions of a long document when one or more of the following happens:

1. New revised versions of packages used.
2. Smaller changes to a fewer number of pages of a long document.
3. No change in the document, but recompiled with revised page numbers as it happens during compilation of journal articles into an issue for printing.
4. Simply you happened to retypeset for no reason and then you're forced to check each page for surprises.

Now you are left with the job of comparing the PDFs generated now and that of previous version and it is not fun. To make the job easier, `pdgpagediff` package is written.

1. Principles

We have a version of document, say, `file1.pdf` and we have its revised version, `file2.pdf`. `pdfpagediff` will create a composite pdf by juxtaposing each page of `file1.pdf` over the corresponding page of `file2.pdf` or vice-versa. Since the pdf's are transparent, you can notice the slightest change visually by simply flipping through the pages.

2. Dependencies

`pdfpagediff` depends on the following packages:

1. `geometry.sty`
2. `graphicx.sty`
3. `color.sty`
4. `substr.sty`

3. Usage

Package can be loaded with the following command:

```
\usepackage{pdfpagediff}
```

Another command `\layerPages` has been defined to include two versions of the PDF documents to create the composite document, the syntax is:

```
\layerPages[<optional page numbers>]{<file1>}{<file2>}
```

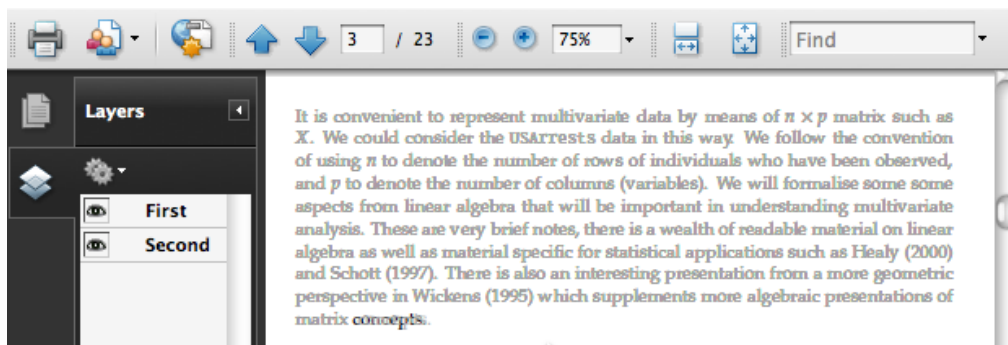
First one doesn't have an optional argument of page numbers, which means all the pages will be used to create the composite document. Second one has comma separated page numbers and hyphen separated page ranges which can be mixed in any order as shown in subsequent examples of usage. The last one has `10-` which means from page `10` to end of the document.

1. `\layerPages{file1.pdf}{file2}`
2. `\layerPages[1,2,4-6,8]{file1}{file2.pdf}`
3. `\layerPages[1,2,4-6,8,10-22]{file1.pdf}{file2.pdf}`

- ▶ Principles
- ▶ Usage
- ▶ Limitations
- ▶ Download
- ▶ Dependencies
- ▶ Examples
- ▶ Acknowledgements

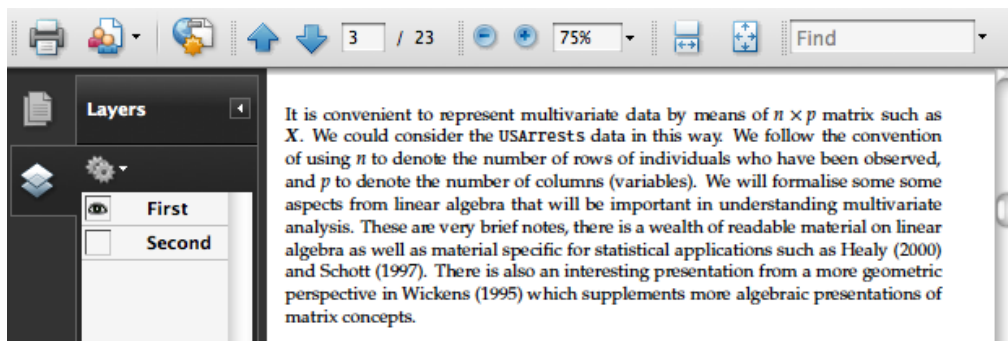
4. `\layerPages[1,2,4-6,8-13,17]{file1}{file2.pdf}`
5. `\layerPages[10-]{file1.pdf}{file2.pdf}`

You need Adobe Reader to view the composite document which only provides to view each layer or all layers together or no layers at all. There is a small layer button at top left hand side of the Adobe Reader window, see the figure below:



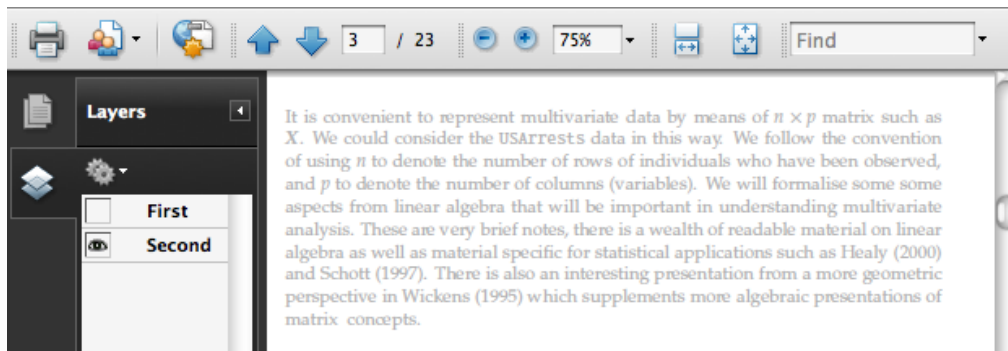
You can see **Layers** icon, clicking on the icon will show you the layers. We have two layers in this example, namely, **First** and **Second** which are also the default. These labels can be changed with `\FirstDoc` and `\SecondDoc` commands respectively.

First Document



The above figure shows the first document alone. You might note that icon for second layer is not visible now.

Second Document



The second document is generated with 30% gray instead of black to facilitate easy indication of locations with differences. Also, note that icon for first layer is invisible since the second layer alone is mode visible here.

4. Examples

You might take a look at Clip 1 which has a paragraph from the composite document. The last word of the paragraph has a mismatch.

Clip 1 A paragraph from composite document., p.3

It is convenient to represent multivariate data by means of $n \times p$ matrix such as X . We could consider the USArrests data in this way. We follow the convention of using n to denote the number of rows of individuals who have been observed, and p to denote the number of columns (variables). We will formalise some some aspects from linear algebra that will be important in understanding multivariate analysis. These are very brief notes, there is a wealth of readable material on linear algebra as well as material specific for statistical applications such as Healy (2000) and Schott (1997). There is also an interesting presentation from a more geometric perspective in Wickens (1995) which supplements more algebraic presentations of matrix concepts.

Now let us take a look at the last two lines of the above para from the first document:

Clip 2 Last two lines from first document., p.3

perspective in Wickens (1995) which supplements more algebraic presentations of matrix concepts.

Here is the same location of the second document:

Clip 3 Last two lines from second document., p.3

perspective in Wickens (1995) which supplements more algebraic presentations of matrix concepts.

The difference is a space added before the last word 'concepts'. If you look at Clip 1 now, you will get to know the difference quickly.

Further differences can be observed in pages 9 and 17 of the included document, **1test.pdf**.

5. Limitations

Following limitations apply:

1. Documents with enormous changes cannot be compared.
2. Documents with opaque backgrounds cannot be compared.
3. Tables and figures with background will not provide any meaningful information even if there are differences.
4. This is not a character by character or word by word diff program, instead it depends largely on your eyes very much.
5. **pdfpagediff** will work only with PDF_TEX and will not work with any other T_EX compilers.

User Manual of pdfpagediff Package

C. V. RADHAKRISHNAN

KEYWORDS: PDF, PDFTEX, L^AT_EX

QUICK LINKS

▶ Principles

▶ Dependencies

▶ Usage

▶ Examples

▶ Limitations

▶ Acknowledgements

▶ Download

6. Acknowledgements

The test document is a chapter namely, `matrices.tex` from a freely available textbook, *Matrix Overview* by Paul Hewson, at: <http://knowledgeforge.net/opentextbook/svn/multivariatestatistics/>. Permission to use this chapter to demonstrate the features of `pdfpagediff` is gratefully acknowledged.

7. Download

The package can be downloaded from <http://www.ctan.org/pkg/pdfpagediff>. Bug reports, feature requests and suggestions can be posted at <http://www.cvr.cc/pdfpagediff/>. The author can be contacted at [<cvr@cvr.cc>](mailto:cvr@cvr.cc).