

Русский RMarkdown для создания pdf-документа

решены проблемы с русским языком при генерации pdf

Илья Кочергин

Аннотация

rmarkdown-документ с русскими символами для корректной генерации pdf через LaTeX

Проблеммы с русским символами и их решение

Источники проблем

- шаблон LaTeX, который используется для генерации PDF-файла не содержит необходимых для поддержки русского языка команд в преамбуле. Русские символы просто исчезают из результирующего документа
- при сохранении графиков в pdf-файл возникают ошибки с русскими символами если для открытия pdf-файла используется функция `pdf("plot_file_name.pdf")`. Русские символы заменяются на непонятные символы в результирующем pdf-файле и вызывают большое количество сообщений типа warning.

Как решать проблему с LaTeX

Добавляем в заголовок документа на языке YAML

```
---
output:
  pdf_document:
    latex_engine: xelatex
keep_tex: true
header-includes:
  - \XeTeXdefaultencoding cp1251
  - \usepackage{xltextra}
  - \usepackage{fontspec}
  - \setmainfont{Times New Roman}
  - \setsansfont{Arial}
  - \setmonofont{Courier New}
  - \newfontfamily{\cyrillicfont}{Times New Roman}
  - \newfontfamily{\cyrillicfonttt}{Courier New}
  - \newfontfamily{\cyrillicfontsf}{Arial}
  - \usepackage[english,russian]{babel}
---
```

Нужно правильно указать кодировку исходного файла:

В Windows часто используется кодировка cp1251:

```
\XeTeXdefaultencoding cp1251
```

Но кодировка utf8 становится все разнообразие символов Unicode.

```
\XeTeXdefaultencoding utf8
```

Как решать проблему с pdf graphical device

При интерактивной работе в R для сохранения графика в pdf-файл мы используем открытие графического устройства таким образом.

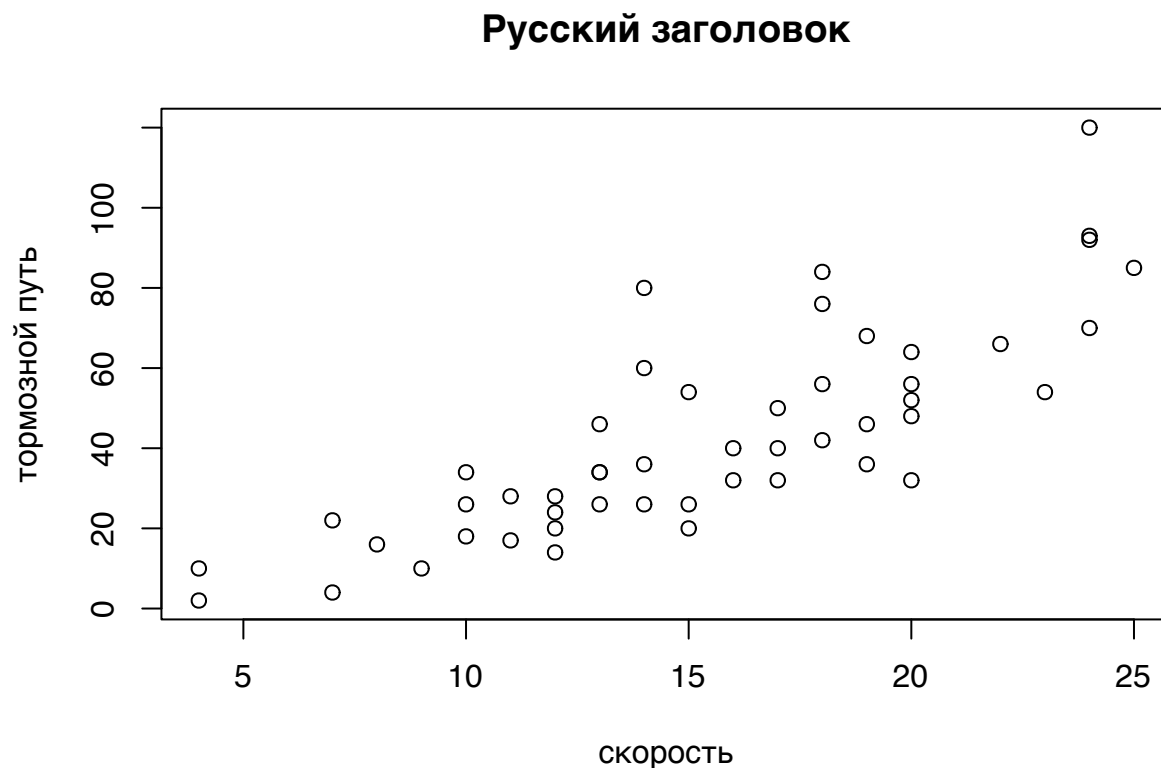
```
pdf("myplot.pdf")
plot(cars, main="Русский заголовок", xlab="скорость", ylab="тормозной путь")
## Warning in title(...): conversion failure on 'Русский заголовок' in
## 'mbcsToSbcs': dot substituted for <d0>
## Warning in title(...): conversion failure on 'Русский заголовок' in
## 'mbcsToSbcs': dot substituted for <a0>
dev.off()
```

Но русские буквы вызывают ошибки.

Для преодоления проблемы достаточно использовать функцию `cairo_pdf()`.

```
cairo_pdf("myplot.pdf")
plot(cars, main="Русский заголовок", xlab="скорость", ylab="тормозной путь")
dev.off()
```

А при опция `dev=chunk'a {r, dev='cairo_pdf'}` решает аналогичную проблему при генерации графиков из Markdown



Этот метод работает не только с базовой графикой, но и с `ggplot2`

Подготовим

```
library(ggplot2)
library(tibble)
heights <- tibble(имя=c("Вася", "Петя", "Коля"),
                  рост=c(172, 188, 162))
```

и нарисует график

```
ggplot(data=heights) +  
  geom_bar(mapping = aes(x=имя, y=рост), stat = "identity")
```

