Applications using R for mapping water quality data in South Africa

Michael Silberbauer

Department of water and Sanitation: SilberbauerM@dws.gov.za

Abstract

Even in a work environment where commercial software licences are available, open source applications can provide flexible and efficient solutions to data visualisation. This has been my experience in developing visual reporting methods for South African water quality monitoring data, mainly using R with packages such as `maptools`, `sp` and `rgdal`. The output finds application in data verification, inventory assessment, document production and web page development.

Keywords

R, water quality, data visualisation, spatial data

1. Introduction

The National Department of Water and Sanitation (DWS) manages an extensive Informix database with close to a million samples from ~60 000 sites. The database record goes back as far as 1970 (Figure 1).
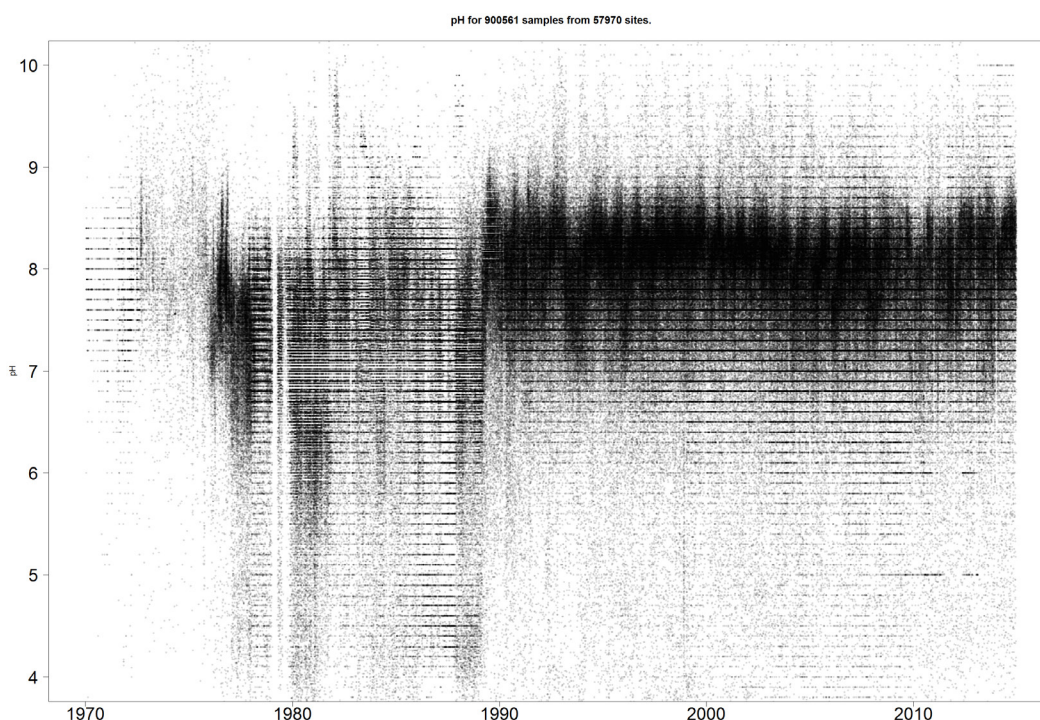

Figure 1. All 900 561 pH results for 57 970 DWS sites from 1970 to 2014. The horizontal striations are a result of rounding. A change in method in 1990 is visible as a difference in variability.

Maps and time series plots are valuable for making sense of such a large mass of data, and as early as 1984 the first water quality map appeared, painstakingly crafted with Letraset™ and Rotring™ (van Vliet & Nell, 1986). Within a few years, the availability of personal computers with software like Turbo Pascal made procedures for graphical presentation of the data much simpler, especially for time series (van Veelen et al., 1990) although the process still involved manual transfer from a mainframe computer to 5.25" disks. When Esri™ products for personal computers became available, scripts to generate specialised map symbols for water quality became an option using primitive graphics (Silberbauer, 2011) but the process has become increasingly difficult as the products matured, and the only built-in multi-variable symbol remains the pie. Simultaneously applying more than one of the 'retinal variables' (e.g. colour and shape; Bertin, 1983/2010) to a graphical element seems only to be possible with complicated manual intervention.

Databases have become much more accessible via personal computers now than in the 1980s. However, accessing the Informix water quality database directly from the commercial software environment is complicated and the results have not been entirely satisfactory.

2. Applications in R

The consequence of the limitations mentioned above is that, whatever the platform used, scripting is necessary for many water quality mapping tasks. The following are examples of using R scripts (R core team, 2015) for reporting chemical, eutrophication and microbial monitoring results on the Internet.

2.1 A simple example to get started

R code for producing Figure 2. In order to use it, one needs shapefiles for catchments and rivers.

```
# SASAqS 2015 example

pts <- data.frame(
  Name=c("RPD",   "SASAqS",   "CT"),
  Lat =c(-25.62, -29.00, -34.00),
  Lon =c( 28.37,  29.47,  18.50),
  Col =c("red", "purple", "blue"),
  Sym =c( 15,      16,      17),
  Val =c( 10.5,   22.0,    12)
)

library(maptools) # (Bivand and Lewin-Koh, 2014)

file.pri <- "C:/data_large/av/drainage/hca_1geo.shp" # polygons
pri <- readShapePoly(file.pri, proj4string=CRS("+proj=longlat +datum=WGS84"))

file.rvr <- "C:/data_large/av/drainage/wriall500_SimplifyLine_500m.shp" # arcs
riv <- readShapeLines(file.rvr, proj4string=CRS("+proj=longlat +datum=WGS84"))

plot(pri, border="grey60", col="white", xlim=c(16, 36), ylim=c(-35, -22), lwd=0.7)

for(ri in 3:7) {
  plot(riv[riv$STRAHLER == ri, ], add=TRUE, lwd=ri/10, col="cyan")
}

points(pts$Lon, pts$Lat, pch=pts$Sym, col=pts$Col, cex=pts$Val / 10)
pointLabel(pts$Lon, pts$Lat, pts$Name, col=pts$Col)

title(main="Using R for mapping water quality data in South Africa")
title(sub=paste("Plotted for SASAqS 2015 on", format(Sys.time(),"%Y-%m-%d %H:%M:%S"),
                "under", R.version.string), cex=0.8, col="grey")

legend(
  "bottomright",
  legend= pts$Name,
  col    = pts$Col,
  pch    = pts$Sym,
  inset = c(0.01, 0.01),
  box.col="grey"
)

box(col="grey")
```
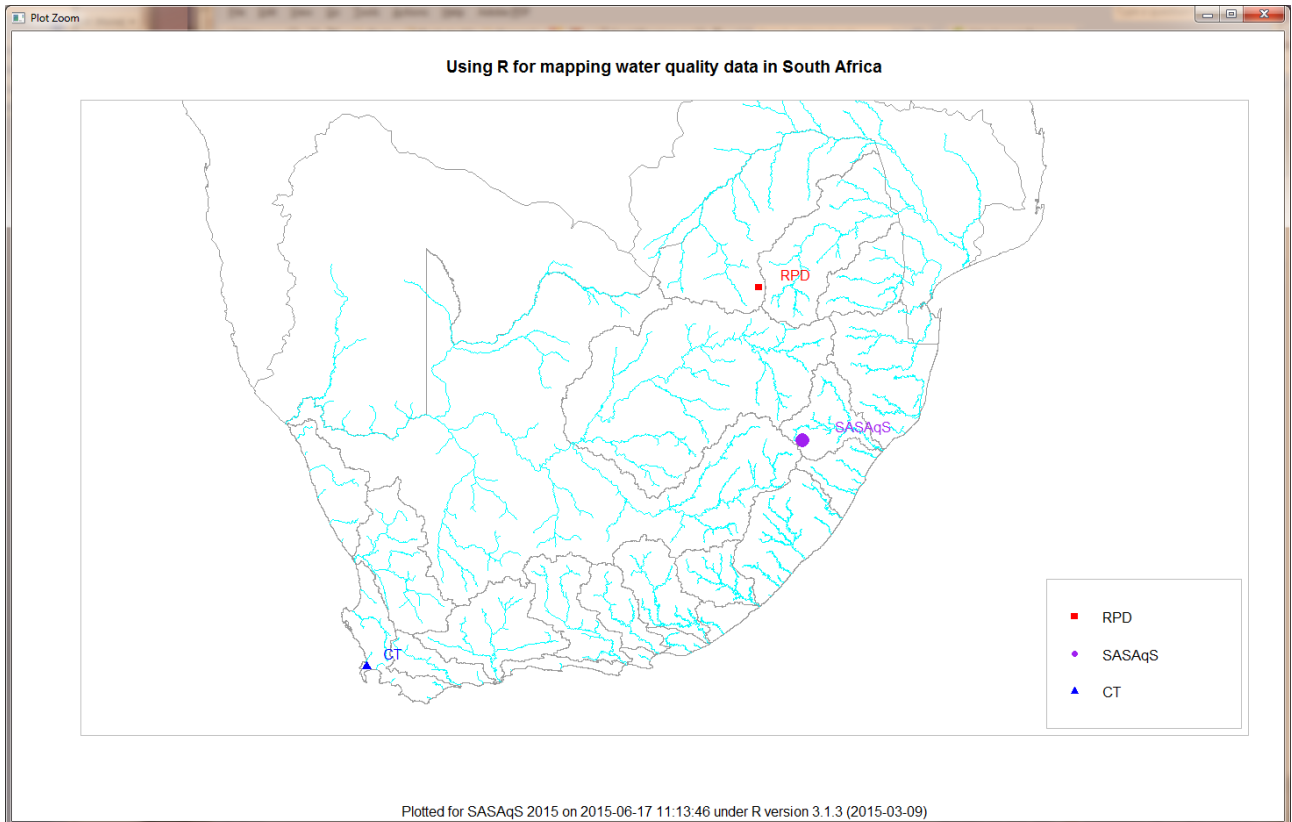
Figure 2. Output from the R `maptools()` example code in the text.

2.2 Chemical water quality data on the web

The next example is not entirely open source, because the output is in Keyhole Markup Language (KML) format and requires a platform like Google Earth to display the output correctly. KML is, however, an Open Geospatial Consortium (OGC) standard.

In 2005, when Google Earth was new, the prototype of this process for generating the KML files and time series summaries ran in Esri's Arc Macro Language (AML) using a subset of the Informix database extracted to the INFO environment (Silberbauer & Geldenhuys, 2008). Conversion to an R-based procedure began in 2011 and comprises three main scripts: shapefile creation from the Informix water quality monitoring site catalogue, KML creation from the shapefile data and plotting of multiple variable time series (Silberbauer, 2011; 2013). The important components in R are:

> Data input using the `RODBC` package via a Windows ODBC link to the Informix database (Ripley and Lapsley 2013).
> The function `writeOGR()` in R package `rgdal` for writing shapefiles (Bivand, Keitt and Rowlingson, 2014).
> The `png()` function with `par(fig=...)` for multiple time series plots of water quality on a single page.

Creation of KML and HTML files is via brute force line-by-line `write()` commands, in order to fine-tune the layout of the output files.

2.3 Eutrophication data on the web

On-the-fly generation of maps introduces security complications within the corporate IT environment, so a compromise solution for eutrophication reporting is the regular preparation of static six-monthly maps for each summer and winter season, combined with an HTML index. Each site has a potential for eutrophication and an actual eutrophication class, requiring binary marker symbols. Furthermore, different marker shapes are required for lentic and lotic environments. The process includes:

> Data input using the `RODBC` package from the Informix database.
> The `maptools` package for map creation.
> The `polygon()` and `plotrix floating.pie()` functions for drawing markers (Lemon, 2006).
> The `png()` function for map image output.

Creation of HTML files is again a brute force line-by-line process using `write()` commands. Additionally, the web page requires isopleths of depth, oxygen and light penetration against time, all of which have their own R scripts.

2.4 Microbial data on the web

Two styles of presentation are currently available on the DWS website: a set of maps for each two-month reporting period with an HTML index, analogous to the eutrophication method, and an interactive map based on leaflet (https://www.dwa.gov.za/iwqs/microbio/nmmp.aspx). The process comprises:

> Site input using the `RODBC` package from the Informix database.
> Placing of sites on a map using `R2leaflet` (Grant, undated).

Clicking on a site opens a link to existing graphics produced in separate microbial data applications in R, namely `NMMP_miniplots.R` and `NMMP_TimeLine.R`.

3. Results and discussion

The KML-based chemical data inventory is a powerful data exploration tool. It works well on Google Earth (Figure 3), but is less informative on Google Maps after recent changes to the map application by Google (Figure 4). The `leaflet` mapping environment is a possible alternative. On a mobile device, thin fingers or a stylus are useful for selecting the correct point in any mapping application.
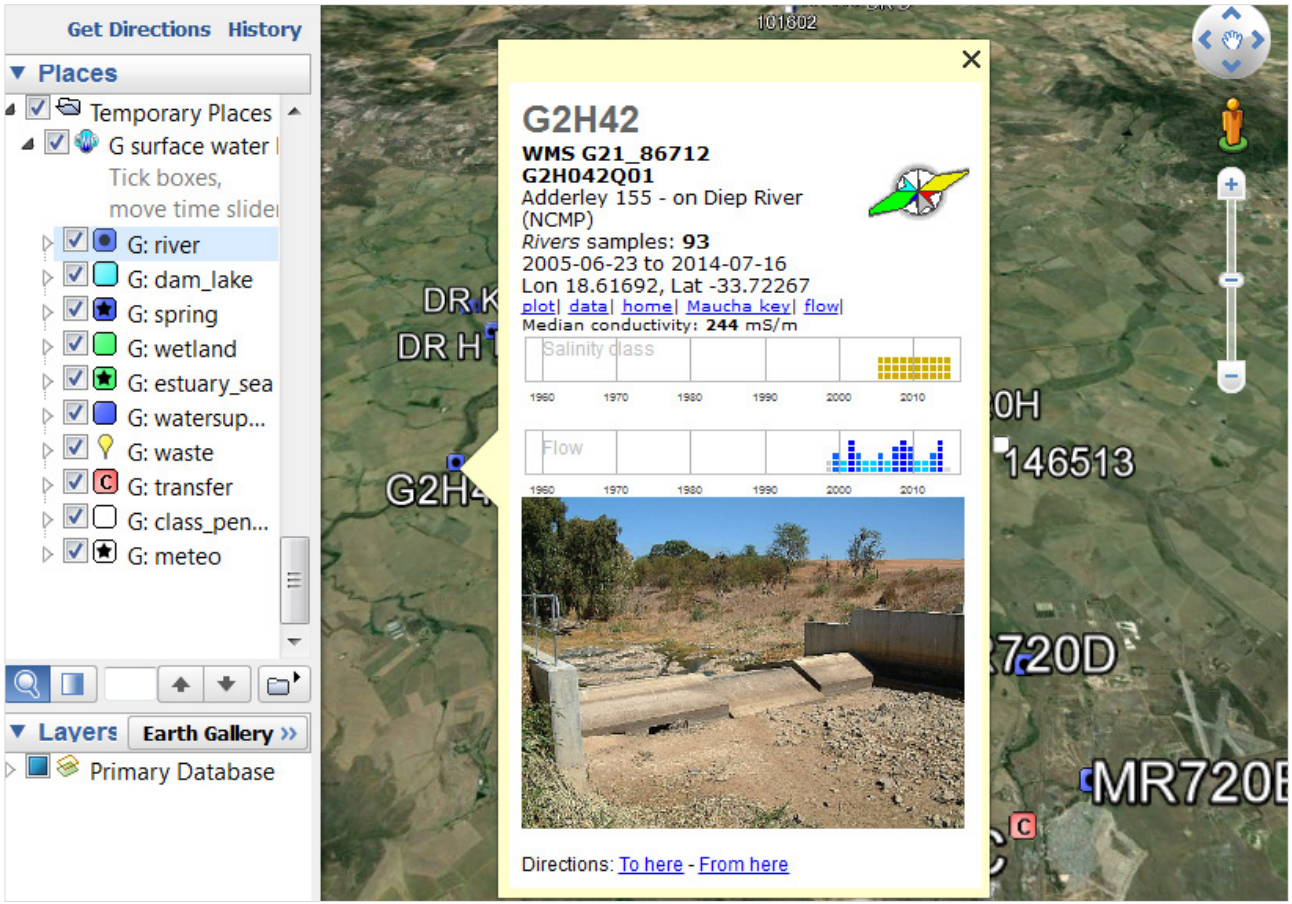
Figure 3. The information available for a monitoring site when the KML output of the chemical reporting script is viewed in Google Earth. https://www.dwa.gov.za/iwqs/wms/data/WMS_pri_txt.asp
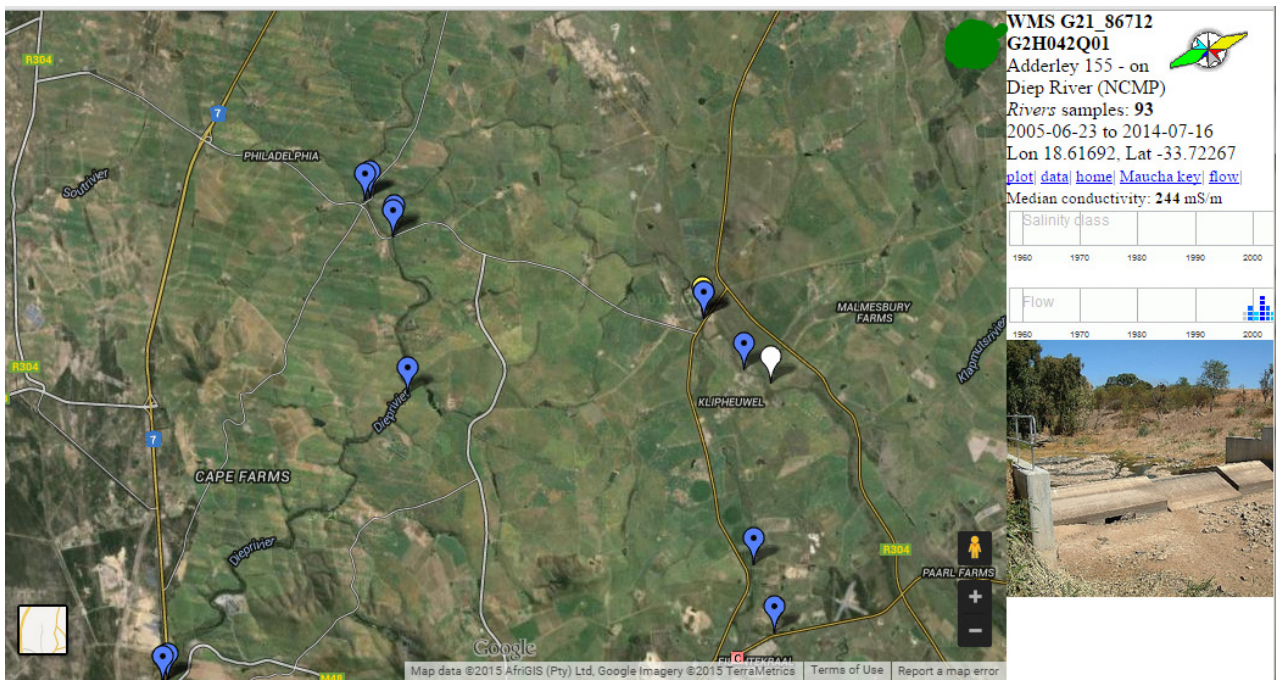

Figure 4. The same site as in the previous figure, viewed in Google Maps.

The eutrophication map set gives an overview of the geographical distribution (Figure 5), while an accompanying table on the website has all the links required for obtaining further information. However, the user interface could do with a great deal of improvement, with an interactive map and direct links from the map, via click or mouseover, to the supporting information.
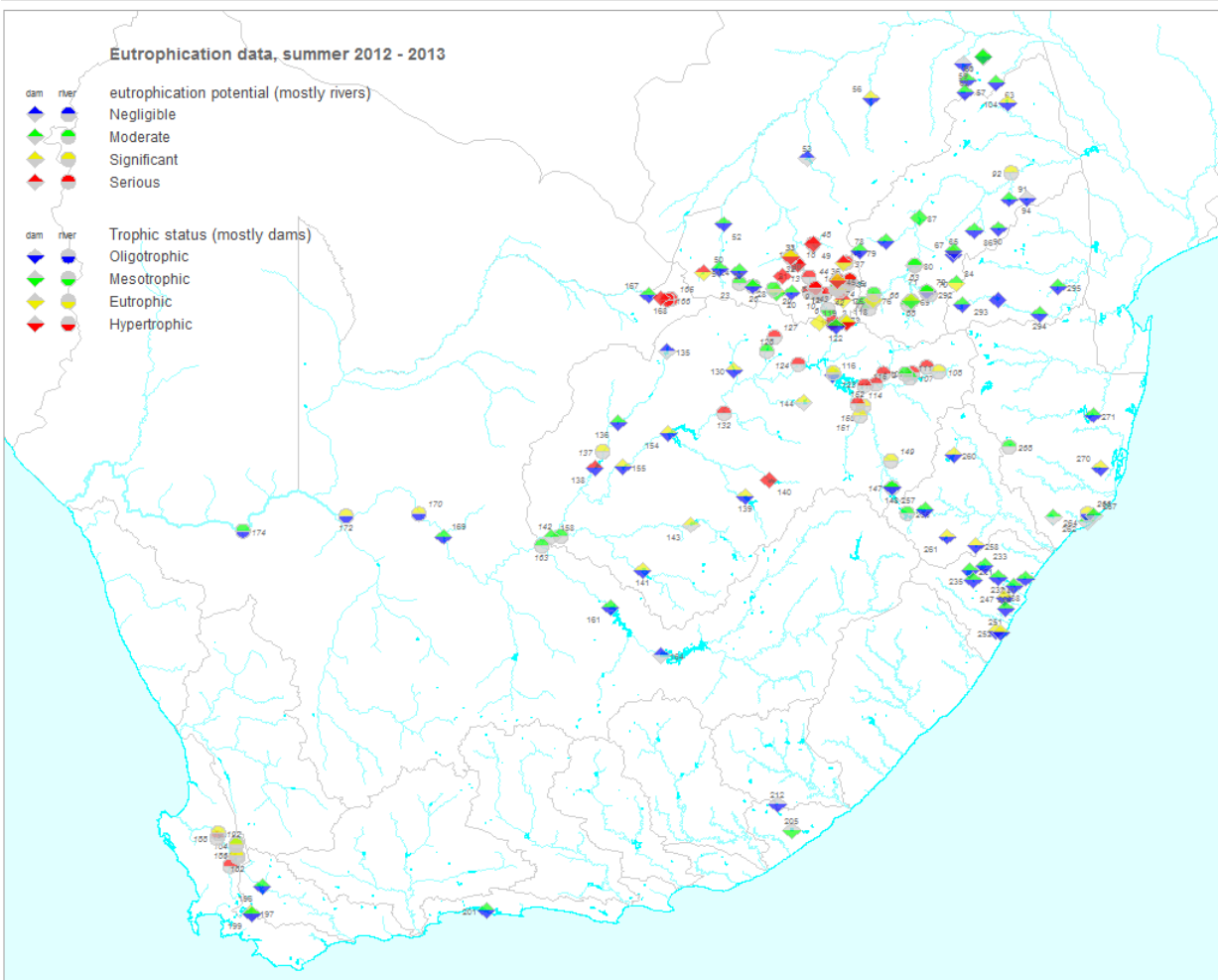
Figure 5. A seasonal eutrophication map for the 2012-2013 summer. The table is too large for inclusion. https://www.dwa.gov.za/iwqs/eutrophication/NEMP/

The microbial map is still experimental, testing the leaflet mapping application as an interface (Figure 6). Most users of microbial data use the standard tables and water management area maps for the information that they need for official reporting (https://www.dwa.gov.za/iwqs/microbio/nmmp.aspx).
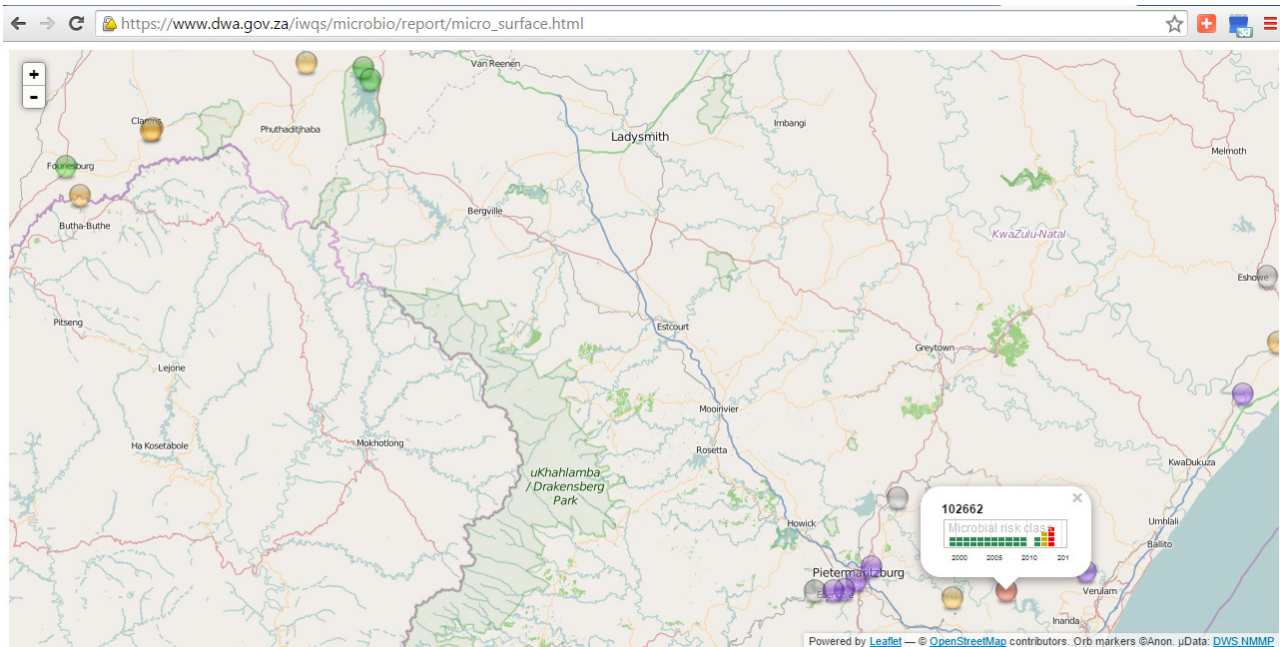
Inventory of microbial monitoring data for surface water

Figure 6. A microbial monitoring map made with the `R2leaflet` script by Grant (undated). https://www.dwa.gov.za/iwqs/microbio/report/micro_surface.html

4. Conclusion

These examples demonstrate that R is more than adequate for water quality data mapping and exploration. However, they only scratch the surface of what is possible with R and with open source software in general. Developments that are pushing the envelope include new data interchange standards for spatial and water data, and comprehensive open source packages such as QGIS, a free geographical information system, and ESA's WOIS, for monitoring water resources using data from remote sensing platforms (QGIS Development Team, 2015; Guzinsky et al., 2014).

References

Bertin, J. (2010). Semiology of Graphics: Diagrams, Networks, Maps. Esri Press, 1st edition.

Bivand, R., Keitt, T. and Rowlingson, B. (2014). `rgdal`: Bindings for the Geospatial Data Abstraction Library. R package version 0.9-1. http://CRAN.R-project.org/package=rgdal

Bivand, R. and Lewin-Koh, N. (2014). `maptools`: Tools for reading and handling spatial objects. R package version 0.8-30. http://CRAN.R-project.org/package=maptools

Grant, R. (undated) Robert Grant Stats: `R2leaflet`. URL http://www.robertgrantstats.co.uk/software.html accessed 2015-03-07.

Guzinski, R., Kass, S., Huber, S., Bauer-Gottwein, P., Jensen, I., Naeimi, V., Doubkova, M., Walli, A., and Tottrup, C. (2014). Enabling the use of earth observation data for integrated water resource management in Africa with the water observation and information system. Remote Sensing, 6(8):7819-7839.

Lemon, J. (2006) `plotrix`: a package in the red light district of R. R-News, 6(4): 8-12. http://cran.r-project.org/doc/Rnews/Rnews_2006-4.pdf

QGIS Development Team (2015). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ripley, B. and Lapsley, M. (2013). `RODBC`: ODBC Database Access. R package version 1.3-10. http://CRAN.R-project.org/package=RODBC

Silberbauer, M. (2011). KML-based online spatial inventory of water chemistry data—R version. Technical Report N/0000/00/RD1/2011, Resource Quality Services, Department of Water Affairs, Pretoria, South Africa.

Silberbauer, M. (2011). Multivariate point data visualisation - geographical information systems developments to aid in water quality management. In Geertman, S., Reinhardt, W., and Toppen, F., editors, The 14th AGILE International Conference on Geographic Information Science- Advancing Geoinformation Science for a Changing World. AGILE.

Silberbauer, M. (2013). BARCODE time-series plots of inorganic water chemistry variables—RODBC. Technical Report N/0000/00/RD2/2011, Resource Quality Information, Department of Water Affairs, Pretoria, South Africa.

Silberbauer, M. J. and Geldenhuys, W. G. (2008). Using Keyhole Markup Language to create a spatial interface to South African water resource data through Google Earth. In Proceedings of the FOSS4G 2008 conference.

van Veelen, M., Nell, U., and Geldenhuys, W. F. (1990). Surface water quality of South Africa 1979-1988. Volume 2: Drainage Region A and B. Technical Report TR145, Hydrological Research Institute, Department of Water Affairs, Pretoria.

van Vliet, H. R. and Nell, U. (1986). Surface water quality of South Africa, the Vaal River catchment 1979-1983. Technical Report TR131, Hydrological Research Institute, Department of Water Affairs and Forestry, Pretoria.