Not just R-spatial: sustaining open source geospatial software stacks

Roger Bivand

2 September 2019, 11:00-13:00, Room 401

Source and links

The presentation files are on github at: (https://github.com/rsbivand/geostat19_talk)

Introduction

- Understanding the original uses of S and R (in the 1980s and 1990s), and seeing how these uses affected the development of R lets us appreciate the robustness of R's ecosystem.
- This talk uses readings of the R sources and other information to explore R's history. The topics to be touched on include the "colour" books (brown, blue, white, green), interlinkages to SICP (Scheme) and LispStat
- This should show where and when impulses flowed between R and the development of capabilities in R for spatial data analysis
- The focus will be on the period 1997-2004, which is when a lot of the foundations for R spatial came into being

History of R and its data structures

Sources

- Rasmus Bååth has a useful blog piece on R's antecedents in the S language
- Something similar is present in the second chapter of Chambers (2016), from the viewpoint of one of those responsible for the development of the S language
- In addition to S, we need to take SICP and Scheme into account (Abelson and Sussman 1996), as described by Ihaka and Gentleman (1996) and Wickham (2014)
- Finally, LispStat and its creators have played and continue to play a major role in developing R (Tierney 1990, 1996, 2005)

R did not always look like an alternative implementation of the S language. It started as a small Scheme-like interpreter (loosely based on work by Sam Kamin [4] and David Betz [2]). This provided a platform for experimentation and extension. The following dialog shows a very early version of the R interpreter at work.

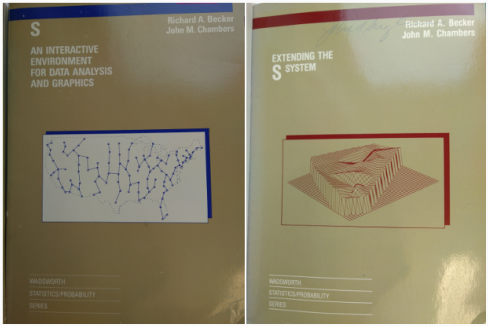
```
> (define square (lambda (x) (* x x)))
square
> (define v 10)
v
> (square v)
100
```

The S-like appearance of R was added incrementally. We initially moved to an S-like syntax for our experiments because, although we were both familiar with Lisp syntax, we didn't feel that it was really suitable for expressing statistical computations. This choice of syntax set us on a path towards compatibility with S. Once initiated, the move towards compatibility with S was irresistible. This was partly because we wanted to see just how far we could push it and partly because it have us access to code resources and developers.

From S to R: Brown Books

Becker and Chambers (1984): S: An Interactive Environment for Data Analysis and Graphics, A.K.A. the Brown Book

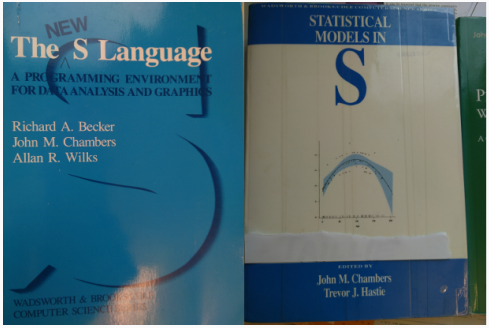
Becker and Chambers (1985): Extending the S System



From S to R: Blue and White Books

Becker, Chambers and Wilks (1988): The New S Language: A Programming Environment for Data Analysis and Graphics, A.K.A. the Blue Book.

Chambers and Hastie (1992): Statistical Models in S, A.K.A. the White Book.



From S to R: Green Book

Chambers (1998): Programming with Data: A Guide to the S Language, A.K.A. the Green Book.

Venables and Ripley (2000): S Programming

S2 to S3 to S4

- The S2 system was described in the Brown Book, S3 in the Blue Book and completed in the White Book, finally S4 in the Green Book
- The big advance from S2 to S3 was that users could write functions; that data.frame objects were defined; that formula objects were defined; and that S3 classes and method dispatch appeared
- S4 brought connections and formal S4 classes, the latter seen in R in the **methods** package (still controversial)
- S-PLUS was/is the commercial implementation of S and its releases drove S3 and S4 changes

S, Bell Labs, S-PLUS

- S was a Bell Labs innovation, like Unix, C, C++, and many interpreted languages (like AWK); many of these share key understandings
- Now owned by Nokia, previously Alcatel-Lucent, Lucent, and AT&T
- Why would a telecoms major (AT&T) pay for fundamental research in computer science and data analysis (not to sell or market other products better)?
- Some Green Book examples are for quality control of telecoms components

S-PLUS and R

- S-PLUS was quickly adopted for teaching and research, and with S3, provided extensibility in the form of libraries
- Most links have died by now, but see this FAQ for a flavour - there was a lively community of applied statisticians during the 1990s
- S built on a long tradition of documentation through examples, with use cases and data sets taken from the applied statistical literature; this let users compare output with methods descriptions
- ... so we get to R

and what about LispStat?

- Luke Tierney was in R core in 1997, and has continued to exert clear influence over development
- Because R uses a Scheme engine, similar to Lisp, under the hood, his insight into issues like the garbage collector, namespaces, byte-compilation, serialization, parallelization, and now ALTREP has been crucial (see also the proposal by Luke Tierney, Gabe Becker and Tomas Kalibera)
- Many of these issues involve the defensive copy on possible change policy involved in lazy evaluation, which may lead to multiple redundant copies of data being present in memory
- Luke Tierney and Brian Ripley have fought hard to let R load fast, something that is crucial to ease the use of R on multicore systems or inside databases

R spatial and R-sig-geo

Antecedents (among others)

- The AI-GEOSTATS listserve/mailing list was started by Gregoire Dubois in 1995; AI was from Arc/Info
- GRASS accepted code contributions, then kept in `src.contrib` and `src.garden` and GRASS mailing lists were active (the archives seem to have been lost in moving from minordomo to mailman in 2001)
- Bao et al. (2000) describe S-PLUS links to GRASS and ArcView; the module for spatial statistics was also from the mid-1990's
- `MASS` included S and compiled code for kriging, trend surfaces and point pattern analysis, `splancs` provided point pattern analysis

Albrecht Gebhardt

- Much of the porting of S code to R for spatial statistics was begun by Albrecht Gebhardt
- The CRAN archives show `tripack` and `akima` - both with ACM licenses - from August 1998; `ash` and `sgeostat` followed in April 1999
- The date for the MASS `spatial` package is unknown because it was bundled in the `VR` metapackage subsequently split into `spatial`, `MASS`, `nnet` and `class`
- In the earliest period, CRAN admins helped practically with porting

- Rowlingson and Diggle (1993) describes the S-PLUS version of `splancs` for point pattern analysis
- I'd contacted Barry in 1997 but only moved forward with porting `splancs` as R's ability to load shared objects advanced
- In September 1998, I wrote to him:

> It wasn't at all difficult to get things running, which I think is a result of your coding - thank you! An issue I have thought about a little is whether at some stage Albrecht and I wouldn't integrate or harmonize the points and pairs objects in Splancs, spatial, and Sgeostat - they aren't the same, but for users maybe ought to appear to be so. Maybe when I get the GRASS interface up?

- The earliest `splancs` in the CRAN archive is from November 2000

1998 ERSA talk

- Albrecht and I were teaching spatial data analysis using R and code that would go into CRAN packages (I still have the package I wrote, and the exams)
- In 1998, the ERSA (European Regional Science Association) meeting was in Vienna, and we gave a talk subsequently published in Bivand and Gebhardt (2000)
- We covered R, point pattern analysis, geostatistics, ESDA and spatial econometrics

tering. It would be of advantage if both Splancs and the `spatial` package defined point objects in the same way, or at least if conversion functions were written, permitting data to be moved between them without uncertainty about the bounding region being used; both initially date from a period when object-oriented mechanisms were only beginning to enter `S`.

- Markus Neteler and I were in touch too, as I'd been trying to interface R and GRASS (R. Bivand 2000)

```
I just uploaded GRASS-0.1-1.tar.gz - a standard R package, and
paper2.pdf.gz - the latest version of the C&G submission, to
hgeo02.geog.uni-hannover.de. The package was tested on 5.0b5 on the GRASS
side, and 0.99.0 (released 7/2) on the R side. In fact, the package caused
a patch to be issued for R yesterday - I was overstressing a strsplit()
function in R in parsing the output of g.list! The same two files are on
ftp://reclus.nhh.no/pub/candg.
```

- The first GRASS package in the CRAN archive is from November 2000
- Before that, we'd presented at GeoComputation 2000 in Chatham Bivand and Neteler (2000)

- Nicholas Lewin-Koh posted on the R list (subsequently split between R-help and R-devel) in September 2000, asking about coding the reading of ESRI shapefiles
- He made a number of key contributions which ended up in `spweights` - currently part of `spdep` - and what became `maptools`
- He wrote the first vector representation based on the shapefile import format, and I should be blamed for subsequent fumbling implementations in `maptools`

DSC 2001

- From interaction with Kurt Hornik, Ross Ihaka and Brian Ripley on the R list during late 2000, contacts widened, including Tim Keitt (`rgdal`)
- I ended up presenting at the DSC meeting in Vienna in March 2001, getting to meet many R core members, developers and CRAN admins

15:15-15:40	Coffee Break	Coffee Break
15:40-17:20	**Applications** Anna Bartkowiak A Set of XLispStat Subroutines for Detecting Outliers Balasubramanian Narasimhan A Java-based test bed for Random Number Generators Chen Xun A Quasi-Exact Method for the Confidence Intervals for the Difference of Two Independent Binomial Proportions in Small Sample Cases Martin Theus What Dataminers Want	**R Connectivity** Roger Bivand R and geographical information systems, especially GRASS Erich Neuwirth Embedding R in standard software, and the other way round Erich Neuwirth R Sings or using R to sonify data Brian Ripley Applications of R clients and servers

- Further contacts came from a CSISS workshop organised by Luc Anselin in May 2002 in Santa Barbara, including Konstantin Krivoruchko (ESRI) and Gilberto Câmara (INPE)
- My talk eventually appeared as Bivand (2006), but the contacts were more useful
- Visiting the INPE earth observation lab and presenting at a conference in December 2002 provided further ideas and links to TerraLib
- Luc himself contributed code to `spdep` and continues be very supportive
- I did a further piece for the ERSA 2002 conference in Dortmund (Bivand 2002)

- The DSC meeting in March 2003 was also being planned for Vienna, and I'd talked about it with Kurt Hornik when he gave a keynote at the GRASS meeting in Trento in September 2002
- I met Virgilio Gómez-Rubio at the GRASS meeting among many others and talked about `DCluster` and `RArcInfo`

10.30-12.30 **Keynote speeches**
Chairman Battista Benciolini, University of Trento, Italy.

Current status and future development of GRASS
Markus Neteler, ITC-irst, Italy.

GRASS as a multidimensional dynamic GIS: past vision and where we are now
Helena Mitasova, Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, USA.

The dynamics/development of a free software project in statistical computing: R
Kurt Hornik, Vienna University of Economics and Business Administration, Austria.

Planning for R spatial

- In mid November I started emailing contacts to invite sumbissions for a spatial statistics session at DSC, and a workshop at which we could discuss working together
- Of the replies, far more would have liked to attend than were able to (Pebesma 2004; Gómez-Rubio, Ferrándiz-Ferragud, and López-Quílez 2005; Gómez-Rubio and López-Quílez 2005)

Friday 21 10:30-12:30	**Spatial Statistics** C. Furlanello, M. Neteler, S. Merler, S. Menegon, S. Fontanari, A. Donini, A. Rizzoli and C. Chemini *GIS and the Random Forest Predictor: Integration in R for Tick-Borne Disease Risk Assessment* V. Gómez Rubio, J. Ferrándiz and A. López *Detecting Disease Clusters with R* Barry Rowlingson, Adrian Baddeley, Rolf Turner and Peter Diggle *Rasp: A Package for Spatial Statistics* Virgilio Gómez Rubio *RArcInfo: Using G.I.S. Data with R*
15:30-17:30	Edzer J. Pebesma *Gstat: Multivariable Geostatistics for S* Paulo J. Ribeiro, Ole F. Christensen and Peter J. Diggle *Geostatistical Software -- geoR and geoRglm* Albrecht Gebhardt *PVM-Kriging with R* E. Glatzer and R. Müller *An Interactive Tool for Residual Diagnostics in Variogram Fitting (with Implementation in R)* Links: Open pre-DSC'03 spatial data workshop Back to Overview

Open pre-DSC'03 spatial data workshop: R, map(), GIS, ...

Wednesday 19 March 2003, Department of Economic Geography and Geoinformatics, Vienna University of Economics and Business Administration

Location: Room RL 1.16, Department of Economic Geography and Geoinformatics, Vienna University of Economics and Business Administration, Rossauer L.�nde 23/1, A-1090 Vienna; access by crossing road from U4 Metro Station Rossauer L.�nde. The Department is at the red ring in the centre of this map.

Times: we expect to run to the following schedule: 0930-1030, coffee, 1100-1230, lunch, 1400-1530, coffee, 1600-end, where "end" dovetails with the DSC icebreaker.

Who: this open workshop is "open" - anyone with an interest in what we will be discussing is welcome to attend, whether you are an active contributor or not.

Aims: the open workshop is intended to permit as many as possible of DSC participants interested in the practical handling of spatial/geographical data to exchange views. During the day, we should be able to hear about various forms of R/GIS integration, both loose and tight coupled. Topics here include packages already on CRAN (RArcInfo, GRASS, and the spatial statistics packages), and initiatives under development, whether active or not (especially Rmap, the R/GDAL package, functions to read and write shapefiles, PostGIS, GRASS 5.1, TerraLib, using StatConnector in ArcGIS, others?).

It would be good if we can exchange and dig up enough information to try to avoid too much duplication of code and work, and to think through some of the issues involved in working with spatial data in particular in the R setting. In some cases, R is a compute engine embedded in the application, in others the spatial data is ingested into R - glancing at Omegahat might also be relevant as a source of ideas.

Facilities: there will be plenty of DHCP network points, power cables (local sockets!) and a beamer all day. If you would like to join lunch at Serviten St.berl, please let me know by 5 March so that we can book a room there. We will make arrangements if you have particular requirements.

Programme: Please let me know if you would like to present or demonstrate your software for linking R and GIS, and/or for accessing and handling spatial data in R. Since some participants will be present for only part of the workshop, please indicate your time constraints. This paragraph will be updated to reflect status.

Contact: Roger Bivand, Economic Geography Section, Department of Economics, Norwegian School of Economics and Business Administration, Breiviksveien 40, N-5045 Bergen, Norway.

- Edzer was already well known because of his open-source `gstat` program (Pebesma and Wesseling 1998)
- From Edzer's reply 13 November 2002

```
I wonder whether I should start writing S classes. I'm afraid I should.
```

Correspondence developed quickly (21 November)

```
> > > I wonder whether I should start writing S classes. I'm afraid I should.
> >
> > Definitely, and probably new-style - have you got Venables & Ripley S
> > programming? Plenty is happening on classes, and they are among the things
> > it would be great to reach a consensus on at the little pre-meeting in
> > Vienna (or by correspondence, but brainstorming many-to-many may be better
> > than multiple one-on-one?
>
> I more or less worked through the green book, by now.

Tim Keitt's interface to GDAL uses new-style classes, so does pixmap now.
I think it will be hard but worthwhile to get a shared set of spatial data
classes built, making it easier for everything on top of them, although I
don't see this as being fun!
```

Contribution even though he couldn't get to Vienna:

> I was looking over all the DSC material, especially the spatial stuff. I did notice, after looking through peoples packages that there is a lot of duplication of effort (I am sure you are aware of this). My suggestion is that we set up a repository for spatial packages similar to the bio-conductor model. Where we have a base spatial package that has S-4 based methods and classes that are efficient and general, Say Spatial_base. Then a Maps, PointProc, LatticeProc, and GeoProc and whatever else packages could be built on top of the base class. i would

E-mail traffic: PROJ4

```
Cc: statsgrass@grass.itc.it
Subject: Re: [STATSGRASS] proj4

On Thu, 8 May 2003, Agustin Lobo wrote:

>
> Hi!
>
> Does anyone know if there is a package to
> call proj4 from R (or python)?
> (proj4: library to make coordinate conversions
> betwee different projections, http://remotesensing.org/proj/)
>

I've been working on one, based on Barry Rowlingson's code from his draft
Rmap package. How soon do you need it, and on what platform? I'm trying to
write class objects for coordinates, projection arguments, etc., so any
early release will be subject to change.
```

```
Date: Thu, 22 May 2003 17:55:56 +0200
From: mailman-admin@stat.math.ethz.ch
To: maechler@stat.math.ethz.ch
Subject: Your new mailing list: r-sig-geo

The mailing list `r-sig-geo' has just been created for you.  The
following is some basic information about your mailing list.

Your mailing list password is
```

StatGIS meeting 2003

- Albrecht Gebhardt was a local organiser for a meeting at which many of those at the DSC workshop met in October 2003

- **Geostatistics: Theory and New Methods**

  (Methods of statistical learning theory, Bayes methods, spatial prediction and risk analysis, spatial sampling and monitoring network design etc)

  Conveners: W. G. Müller, J. Pilz

- **Combining Statistics and GIS**

  (Interfaces between (geo-)statistics and GIS, integration of geostatistics and GIS, mutual benefits, statistical problems of error propagation and uncertainty in GIS, etc)

  Conveners: R. Bivand, G. Dubois

- **Geostatistical Applications**

  (Application of geostatistical methods in the Earth and Environmental Sciences, in Agriculture and Forestry, Epidemiology and Health Sciences, Econometrics and Telecommunications)

  Conveners: H. Glass, M. Kanevsky

- **Spatial Data Bases and Mapping**

  (Interfaces between (geo-)statistical software systems, spatial database management systems and visualization and mapping software systems)

  Conveners: V. Gomez, B. Rowlingson

- **Geostatistical Software Developments**

  (New implementations of geostatistical methodology in R, S-Plus, SAS and other statistical software systems)

  Conveners: E. Pebesma, P. Ribeiro

- R News class description: (Pebesma and Bivand 2005)

Friday 21 16:45-18:15	**Spatial Statistics** Roger Bivand, Edzer Pebesma and Barry Rowlingson *Generic functions for spatial data* Albrecht Gebhardt and Claudia Gebhardt *Bayesian Methods in Geostatistics: Using prior knowledge about trend* Nadine Henkenjohann *spatialreg - A package to fit spatial regression models with isotropic and anisotropic covariance functions* Juergen Pilz and Philipp Pluch and Gunter Spoeck *Taking into account uncertainty in spatial covariance estimation for Bayesian prediction* Back to Overview | Session Format

CRAN arrivals

- `fields` and `geoR` in mid-2001, `spatstat` in early 2002
- `gstat` in February 2003, `maptools` in August 2003, `rgdal` in November 2003
- `sp` in April 2005 (after a workshop in Lancaster November 2004 and before one in Valencia May 2005)
- `classInt` in March 2006, `spgrass6` in August 2006

Tutorial: Analysing Spatial Data in R

presented by Roger Bivand

Goals:

> The workshop will introduce the foundation classes for spatial data provided in the sp package, and show how they make work with spatial data easier. The classes have shared interfaces to common file formats for spatial data, and a range of subsetting, plot, overlay, and sampling methods.
>
> The workshop will let participants needing to import spatial data into R, analyse it in one or more packages, display it in graphical form, and perhaps export it from R, learn how these steps can be made conveniently and robustly. The final section will show how the new class structure makes writing custom interfaces less challenging, making R analysis functions available for a wide range of spatial data formats.

Programme:

1. Representing spatial data in R (30 minutes)
2. Vizualising spatial data in R (30 minutes)
3. Accessing spatial data in R (30 minutes)
4. break
5. Worked examples: geostatistics (30 minutes)
6. Worked examples: disease mapping (30 minutes)
7. Worked examples: customising spatial data classes (30 minutes)

Workshop materials:

> pre-workshop notes, scripts, and data sets on Rgeo website
> download and install at least the core packages from the Spatial task view.

Post-workshop follow-up on the R-sig-geo mailing list.

Back to Tutorials

Thursday 15 15:00-18:30	**Ecology & Spatial Statistics** (Spotlights: HS 0.5, Forum: Atrium) Philippe Grosjean, Richard Hillary, Ernesto Jardim, Laurie Kell, Iago Mosqueira, Jan Jaap Poos, Robert Scott and Hunter S. Thompson *Fisheries modelling in R: The FLR (Fisheries Library in R) project* Thomas Petzoldt, Karsten Rinke and Louis Kates *Population ecology modelling with R: A comparison of object oriented approaches* Daniel Doktor *Spatial and statistical modelling of phenological data* T. Laurent, A. Ruiz-Gazen, and C. Thomas-Agnan *GEOXP: An R package for interactive exploratory spatial data analysis* Vojtech Janousek, Vojtech Erban and Colin Farrow *Using the R language for graphical presentation and interpretation of compositional data in mineralogy: Introducing the package GCDkit-Mineral* Rudolf Dutter *Data Analysis System with Graphical Interface* Norbert Solymosi, Andrea Harnos, Jenő Reiczigel and Ferenc Speiser *RpostGIS an R-library for using PostGIS spatial structures and functions* A. Pedro Duarte Silva, Jorge Cadima, Manuel Minhoto and Jorge Orestes Cerdeira *Subselect0.99: Selecting variable subsets in multivariate linear models* Lisbeth Riis and Mikkel Grum *Using R to Reduce Pesticide Usage in the Horticultural Industry* Keiji Osaki *Spatial characteristics of vegetation index map in urban area derived by variogram analysis* Back to Overview | Session Format

Open Source Geospatial software stacks

Software component stacks

- A chapter written in 2011 and published three years later has an overview (Bivand 2014)
- Before discussing software component stacks for GC, we should acknowledge the importance of open standards for geospatial data interchange
- Kralidis (2008) points out the importance of concepts such as that of spatial data infrastructure, whether established within national jurisdictions, within supranational jurisdictions, or by international standards organisations
- A fuller coverage of the relationships between spatial data infrastructures and free and open source geospatial software is given by Steiniger and Hunter (2012)
- Kralidis also helpfully distinguishes between formal, de facto, and ad hoc standards, which provide the flexibility needed to move ahead somewhat faster than standards committees are usually able to do

Component stacks

- Software components appear to have been defined first by McIlroy (1969), as interchangeable subassemblies by analogy with mass production manufacturing
- The software component stack has been a core concept of programming at least since the publication of Kernighan and Plauger(1976), systematising the experience of Bell Labs computer scientists
- Some of the lessons are made clear in programming itself (Kernighan and Pike 1999), while others affect how one may “glue” small utility functions together in an interactive and/or scripting language (Kernighan and Pike 1984)
- Consequently, a software component stack can be taken as sequence of component programs that are used together to achieve a common goal

Component stacks

- Using stacks of components becomes attractive when task objectives can more easily be met by using components developed by others than by developing them independently
- When the costs of keeping a stack working exceed those of rewriting, the stack may fail, but this is seldom the case
- Open source software developers often advertise application programming interfaces (API), with an implicit promise that other downstream developers using the API will be less subject to incompatible changes
- It is then vital that changes in these underlying components do not change the way that dependent components function, unless their earlier behaviour had been in error

Dependency challenges

- As already noted, developers wishing to integrate software components in stacks must pay careful attention to the versioning of the components, and to the impacts of upstream changes on downstream components
- The terms upstream and downstream refer to the ordering of the components, with data flowing from upstream to downstream components
- If the specification of an upstream component changes, those following it will need to be modified
- If the changes are forced by real bugs being fixed, or security holes being blocked, downstream components must react in appropriate ways

Dependency challenges

- However, some changes occur for other reasons, such as code cleaning, reimplementation, or the resolution of licence issues in otherwise functioning code
- In most cases, upstream developers then attempt to reduce changes in their interfaces with downstream components to an unavoidable minimum
- Open source projects are typically most constrained with respect to developer time for maintenance, including the revision of functioning code to accommodate upstream changes that may not improve downstream performance
- A particularly troublesome issue for dynamically linked software components in relatively long-running applications is that of thread safety

Open source geospatial projects

- The Open Source Geospatial Foundation was brought into being in 2006 as a successor to the Mapserver Foundation, itself created the year before
- In addition to providing a shared infrastructure and procedural framework for web mapping, desktop application and geospatial library projects, OSGeo aims to promote open source geospatial software use and development, including use integrated with proprietary software
- McIhagga (2008) discusses some of the ways in which communities of practice have developed, with particular reference to web mapping, and in his description, the open source web mapping “ecology”
- The Geospatial Data Abstraction Library is a crucial part of the upstream geospatial library infrastructure

- One of the most important components required by geospatial applications is the provision of robust and clear representations of coordinate reference systems
- Because GEOS uses OGC SFS specifications for geometries, it does not “build” topologies in the classical GIS arc-node understanding
- GRASS (Geographic Resources Analysis Support System) was already twenty years old when the GRASS developers collaborated in founding OSGeo, and they have been playing an important role in the broader OSGeo movement (Neteler et al. 2008); many of the more recent developments in GRASS are covered by Neteler et al. (n.d.)

Spatial reference systems

Spatial data typically combine position data in 2D (or 3D), attribute data and metadata related to the position data. Much spatial data could be called map data or GIS data. We collect and handle much more position data since global navigation satellite system (GNSS) like GPS came on stream 20 years ago, earth observation satellites have been providing data for longer.

```
> library(osmdata)
> library(sf)
> bbox <- opq(bbox = 'bergen norway')
> byb0 <- osmdata_sf(add_osm_feature(bbox, key = 'railway',
+   value = 'light_rail'))$osm_lines
> tram <- osmdata_sf(add_osm_feature(bbox, key = 'railway',
+   value = 'tram'))$osm_lines
> byb1 <- tram[!is.na(tram$name),]
> o <- intersect(names(byb0), names(byb1))
> byb <- rbind(byb0[,o], byb1[,o])
```

Baseline WKT and PROJ4

Spatial reference systems define how the geoid is viewed (prime meridian, ellipsoid, datum), and, if projected to the plane, where we are (central longitude, latitude, offsets, etc.). Projection (no datum change) and transformation are possible using PROJ and its legacy `proj_api.h` interface directly (`rgdal::spTransform()` and `lwgeom::st_transform_proj()`), or through GDAL (`sf::st_transform()`).

```
> (WKT <- st_crs(byb))
## Coordinate Reference System:
##   EPSG: 4326
##   proj4string: "+proj=longlat +datum=WGS84 +no_defs"
> strwrap(gsub(",", ", ", st_as_text(WKT)))
## [1] "GEOGCS[\"unknown\", DATUM[\"WGS_1984\", SPHEROID[\"WGS 84\", 637813
## [2] "298.257223563, AUTHORITY[\"EPSG\", \"7030\"]], AUTHORITY[\"EPSG\", "
## [3] "PRIMEM[\"Greenwich\", 0, AUTHORITY[\"EPSG\", \"8901\"]], UNIT[\"deg
## [4] "0.0174532925199433, AUTHORITY[\"EPSG\", \"9122\"]], AXIS[\"Longitud
## [5] "EAST], AXIS[\"Latitude\", NORTH]]"
> byb_utm <- st_transform(byb, crs=32632)
> st_crs(byb_utm)
## Coordinate Reference System:
##   EPSG: 32632
##  proj4string: "+proj=utm +zone=32 +datum=WGS84 +units=m +no_defs"
```

Big bump coming:

‘Fast-forward 35 years and PROJ.4 is everywhere: It provides coordinate handling for almost every geospatial program, open or closed source. Today, we see a drastical increase in the need for high accuracy GNSS coordinate handling, especially in the agricultural and construction engineering sectors. This need for geodetic-accuracy transformations is not satisfied by “classic PROJ.4”. But with the ubiquity of PROJ.4, we can provide these transformations “everywhere”, just by implementing them as part of PROJ.4’ (Evers and Knudsen 2017).

Escaping the WGS84 hub: PROJ and OGC WKT2

Following the introduction of geodetic modules and pipelines in PROJ 5 (Knudsen and Evers 2017; Evers and Knudsen 2017), PROJ 6 moves further. Changes in the legacy PROJ representation and WGS84 transformation hub have been coordinated through the GDAL barn raising initiative. Crucially WGS84 often ceases to be the pivot for moving between datums. A new OGC WKT is coming, and an SQLite EPSG file database has replaced CSV files. SRS will begin to support 3D by default, adding time too as SRS change. See also PROJ migration notes

```
> head(st_coordinates(st_geometry(byb)[[1]]), n=1)
##                   X        Y L1
## 4870663682 5.333375 60.30436  1
> system(paste("echo 5.333375 60.30436 |",
+   "proj +proj=pipeline +ellps=WGS84",
+   "+step +init=epsg:32632"), intern=TRUE)
## Warning in system(paste("echo 5.333375 60.30436 |", "proj +proj=pipeline
## +ellps=WGS84", : running command 'echo 5.333375 60.30436 | proj +proj=pipe
## +ellps=WGS84 +step +init=epsg:32632' had status 3
## character(0)
## attr(,"status")
## [1] 3
> head(st_coordinates(st_geometry(byb_utm)[[1]]), n=1)
##             X       Y L1
## [1,] 297436.1 6690941  1
```

CRAN packages

CRAN

- Once S3 permitted extension by writing functions, and packaging functions in libraries, S and R ceased to be monolithic
- In R, a library is where packages are kept, distinguishing between base and recommended packages distributed with R, and contributed packages
- Contributed packages can be installed from CRAN (infrastructure built on CPAN and CTAN for Perl and Tex), Bioconductor, other package repositories, and other sources such as github
- With over 14500 contributed packages, CRAN is central to the R community, but is stressed by dependency issues (CRAN is not run by R core)

CRAN package clusters

- Andrie de Vries Finding clusters of CRAN packages using igraph looked at CRAN package clusters from a page rank graph
- Here we will be choosing the packages listed in the Spatial and SpatioTemporal task views as contained in the **ctv** package, and their pagerank relationships

CRAN package page rank scores

```
##         Rcpp           sp         MASS        knitr     maptools      lattice
##     0.025096     0.019013     0.011718     0.009346     0.008304     0.008235
##       raster         rgdal     testthat        spdep           sf       Matrix
##     0.008017     0.007942     0.007726     0.007601     0.007491     0.007322
##     spatstat surveillance        rgeos      ggplot2          plm         lme4
##     0.007242     0.006626     0.006332     0.006198     0.005794     0.005361
##      plotKML RColorBrewer    rmarkdown     magrittr          rgl      stplanr
##     0.005194     0.005117     0.005046     0.004971     0.004775     0.004681
## RandomFields         gstat      pkgdown     jsonlite        spaMM    spacetime
##     0.004609     0.004561     0.004443     0.004396     0.004271     0.004232
```

First package cluster

```
##           sp        knitr     maptools      lattice       raster        rgdal
##  0.019012564  0.009345800  0.008304410  0.008235211  0.008017250  0.007942472
##     testthat           sf        rgeos      ggplot2      plotKML RColorBrewer
##  0.007726481  0.007491436  0.006332165  0.006198257  0.005193832  0.005117450
##    rmarkdown      stplanr        gstat    spacetime      stringr        Hmisc
##  0.005045938  0.004681475  0.004561265  0.004231588  0.003867439  0.003660129
##         ctmm       igraph
##  0.003615484  0.003579965
```

Second package cluster

```
##         MASS        spdep       Matrix     spatstat surveillance RandomFields
##  0.011717973  0.007600548  0.007322455  0.007241578  0.006626093  0.004608841
##         maps         nlme       fields      foreach     CARBayes    geostatsp
##  0.003930653  0.003588663  0.003572694  0.003291278  0.002865385  0.002855322
##      FLightR          car   CARBayesST         mgcv       georob         boot
##  0.002855058  0.002713727  0.002649398  0.002583666  0.002525372  0.002430476
##       marmap         lgcp
##  0.002401935  0.002258385
```

Third package cluster

```
##    magrittr     pkgdown    jsonlite     mapview          R6        httr
## 0.004970901 0.004443446 0.004395570 0.003621532 0.003488510 0.003479149
##      digest     leaflet       shiny   htmltools     usethis   geojsonio
## 0.003275301 0.003153439 0.003071180 0.003012599 0.002779159 0.002386978
##    roxygen2 htmlwidgets        covr        curl       lintr     geojson
## 0.002361102 0.002127728 0.002106620 0.001987502 0.001905504 0.001880350
##        xml2      RNeXML
## 0.001817758 0.001780216
```

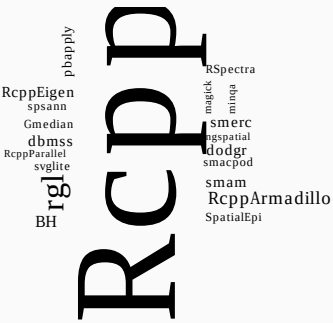
Fourth package cluster

```
##          Rcpp           rgl RcppArmadillo            BH     RcppEigen
##  0.0250960963  0.0047750208  0.0021832309  0.0020183254  0.0017848395
##         dbmss         dodgr          smam         smerc       pbapply
##  0.0017440194  0.0017094478  0.0016755278  0.0016261759  0.0014529893
##      RSpectra    SpatialEpi       smacpod       Gmedian       svglite
##  0.0012778071  0.0011703368  0.0011503814  0.0011358369  0.0010588067
##        spsann         minqa     ngspatial  RcppParallel        magick
##  0.0010420391  0.0010214454  0.0010032247  0.0009765152  0.0009725605
```


maptools
RColorBrewer
raster
testthat
ggplot2
stringr
sf
spacetime
knitr
Hmisc
gstat
igraph
rgeos
rgdal
ctmm
sp
stplanr
lattice
rmarkdown
plotKML


Matrix
RandomFields
spdep
boot
surveillance
spatstat
maps
geostatsp
marmap
car
lgcpgeorob
nlme
mgcv
fields
FLightR
CARBayesST
CARBayes
foreach
MASS

CRAN third and fourth page rank clusters



Roundup: history

- R itself is very varied in the way that approaches have established niches through time
- sp classes date back to the same time as Bioconductor, which has a fairly definite S4 preference
- Communities develop and change over time, and continue to do so, with simultaneous differences in preferences
- so what comes after tidyverse and “data science”? Should we be preparing already?

Abelson, Harold, and Gerald Jay Sussman. 1996. *Structure and Interpretation of Computer Programs*. Boston, MA: MIT Press.

Bao, Shuming, Luc Anselin, Doug Martin, and Diana Stralberg. 2000. “Seamless Integration of Spatial Statistics and GIS: The S-PLUS for ArcView and the S+Grassland Links.” *Journal of Geographical Systems* 2 (2): 287–306.

Becker, R.A., and J.M. Chambers. 1984. *S: An Interactive Environment for Data Analysis and Graphics*. Pacific Grove, CA, USA: Wadsworth & Brooks/Cole.

———. 1985. *Extending the S System*. Pacific Grove, CA, USA: Wadsworth & Brooks/Cole.

Becker, Richard A., John M. Chambers, and Allan R. Wilks. 1988. *The New S Language*. London: Chapman & Hall.

Bivand, Roger. 2000. “Using the R Statistical Data Analysis Language on GRASS 5.0 GIS Database Files.” *Computers & Geosciences* 26 (9): 1043–52. https://doi.org/https://doi.org/10.1016/S0098-3004(00)00057-1.

———. 2002. “Spatial Econometrics Functions in R: Classes and Methods.” *Journal of Geographical Systems* 4 (4): 405–21. https://doi.org/10.1007/s101090300096.

———. 2006. “Implementing Spatial Data Analysis Software Tools in R.” *Geographical Analysis* 38 (1): 23–40. https://doi.org/10.1111/j.0016-7363.2005.00672.x.

———. 2014. “Geocomputation and Open Source Software:Components and Software Stacks.” In *Geocomputation*, edited by Robert J. Abrahart and Linda M. See, 329–55. Boca Raton: CRC Press. http://hdl.handle.net/11250/163358.

Bivand, Roger, and Albrecht Gebhardt. 2000. “Implementing Functions for Spatial Statistical Analysis Using the R Language.” *Journal of Geographical Systems* 2 (3): 307–17. https://doi.org/10.1007/PL00011460.

Bivand, Roger, and Markus Neteler. 2000. *Open Source Geocomputation: Using the R Data Analysis Language Integrated with GRASS GIS and PostgreSQL Data Base Systems*. http://www.geocomputation.org/2000/GC009/Gc009.htm.

Chambers, John M. 1998. *Programming with Data*. New York: Springer.

———. 2016. *Extending R*. Boca Raton: Chapman & Hall.

Chambers, John M., and Trevor J. Hastie. 1992. *Statistical Models in S*. London: Chapman & Hall.

Evers, Kristian, and Thomas Knudsen. 2017. *Transformation Pipelines for Proj.4*. https://www.fig.net/resources/proceedings/fig_proceedings/fig2017/papers/iss6b/ISS6B_evers_knudsen_9156.pdf.

Gómez-Rubio, V., J. Ferrándiz-Ferragud, and A. López-Quílez. 2005. “Detecting Clusters of Disease with R.” *Journal of Geographical Systems* 7: 189–206.

Gómez-Rubio, V., and A. López-Quílez. 2005. “RArcInfo: Using GIS Data with R.” *Computers and Geosciences* 31: 1000–1006.

Ihaka, Ross, and Robert Gentleman. 1996. “R: A Language for Data Analysis and Graphics.” *Journal of Computational and Graphical Statistics* 5 (3): 299–314. https://doi.org/10.1080/10618600.1996.10474713.

Kernighan, Brian W., and Rob Pike. 1984. *The UNIX Programming Environment*. Englewood Cliffs, N. J.: Prentice-Hall.

———. 1999. *The Practice of Programming*. Reading, Mass.: Addison-Wesley.

Kernighan, Brian W., and P. J. Plauger. 1976. *Software Tools*. Reading, Mass.: Addison-Wesley.

Knudsen, Thomas, and Kristian Evers. 2017. *Transformation Pipelines for Proj.4*. https://meetingorganizer.copernicus.org/EGU2017/EGU2017-8050.pdf.

Kralidis, Athanasios Tom. 2008. “Geospatial Open Source and Open Standards Convergences.” In *Open Source Approaches in Spatial Data Handling*, edited by G. B. Hall and M. Leahy, 1–20. Berlin: Springer-Verlag.

McIhagga, David. 2008. “Communities of Practice and the Business of Open Sourceweb Mapping.” In *Open Source Approaches in Spatial Data Handling*, edited by G. B. Hall and M. Leahy, 49–64. Berlin: Springer-Verlag.

McIlroy, M. D. 1969. “‘Mass produced’ software components.” In *Software Engineering*, edited by Peter Naur and Brian Randell, 79–87. Brussels: Conference on Software Engineering, NATO Science Committee; NATO Scientific Affairs Division.

Neteler, Markus, M. Hamish Bowman, Martin Landa, and Markus Metz. n.d. “GRASS GIS: A multi-purpose open source GIS.” *Environmental Modelling & Software* 31 (1): 124–30.

Neteler, M., D.E. Beaudette, P. Cavallini, L. Lami, and J. Cepicky. 2008. “GRASS GIS.” In *Open Source Approaches in Spatial Data Handling*, edited by G. B. Hall and M. Leahy, 171–99. Berlin: Springer-Verlag.

Pebesma, Edzer J. 2004. “Multivariable Geostatistics in S: The Gstat Package.” *Computers and Geosciences* 30: 683–91.

Pebesma, E. J., and R. S. Bivand. 2005. “Classes and Methods for Spatial Data in R.” *R News* 5 (2): 9–13.

Pebesma, E. J., and C. G. Wesseling. 1998. “Gstat, a Program for Geostatistical Modelling, Prediction and Simulation.” *Computers and Geosciences* 24: 17–31.

Rowlingson, B., and P. J. Diggle. 1993. “Splancs: Spatial Point Pattern Analysis Code in S-Plus.” *Computers and Geosciences* 19: 627–55.

Steiniger, Stefan, and Andrew J. S. Hunter. 2012. “Free and Open Source GIS Software for Building a Spatial Data Infrastructure.” In *Geospatial Free and Open Source Software in the 21st Century*, edited by Erwan Bocher and Markus Neteler, 247–61. Lecture Notes in Geoinformation and Cartography. Berlin Heidelberg: Springer.

Tierney, Luke. 1990. *LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*. Wiley, New York, NY.

———. 1996. “Recent Developments and Future Directions in Lisp-Stat.” *Journal of Computational and Graphical Statistics* 5 (3): 250–62.

———. 2005. “Some Notes on the Past and Future of Lisp-Stat.” *Journal of Statistical Software* 13 (9): 1–15. https://doi.org/10.18637/jss.v013.i09.

Venables, William N., and Brian D. Ripley. 2000. *S Programming*. New York: Springer. http://www.stats.ox.ac.uk/pub/MASS3/Sprog/.

Wickham, Hadley. 2014. *Advanced R*. Boca Raton, FL: Chapman & Hall. http://adv-r.had.co.nz/.