This script will be the “official” to extract the data from the imagery in our department. This needs to be robust, flexible, easy to (re)use. In this document I’m trying to order the ideas, requirements and options for this script.

# Inputs

* Option to create a buffer from the given spatial points (eg. shp) to use as zones to extract the values from the imagery.
* Batch options:
  + Extract all images in folder – one zone vector.
    - Many dates-one-trait
    - Many traits-one-date
  + Manage many fields
  + Extract images, give the list of images paths and corresponding zone vectors.
    - Generate a “list of inputs” generator
      * Make sure to warn for unmatching names of img and vector
      * Check if file exists
  + Extract giving a list of what to extract and assume the “greensplane” folder structure.
  + Multiband imagery
    - How to get band name?
* Standard input name structure
* Extract using vectors from a shp or GDB

# Outputs

* Descriptive statistic tables for each band
* Aggregate by Field, date, project…
* Calculate indexes
* Quality control: histograms, outliers, repetitions, saturation…
* Save statistics in the vector files?
* Apply filters?

# Options of Language/tools to use

## Python

I have used this because of ArcGIS, but I want it to be open source. Python looks cleaner in the code.

### Rasterstats Package

### <https://pypi.org/project/rasterstats/>

* Mentioned by Paul J in <https://stackoverflow.com/questions/5802649/zonal-statistics-qgis>
* Used by Alex Rothenberg in <https://sites.google.com/site/alexrothenberg/programs>

### zonal\_stats.py

### <https://gist.github.com/perrygeo/5667173>

* By PerryGeo
* Python implementation of zonal statistics function. Optimized for dense polygon layers, uses numpy, GDAL and OGR to rival the speed of starspan.
* Some comments mention memory issues on large datasets

### SAGA

### <http://www.saga-gis.org/saga_api_python/index.html>

## R

I have more practice in R

### Extract::raster

* Uses vector and image
* Can do some statistics
* Takes time

### Velox Package

<https://www.rdocumentation.org/packages/velox/versions/0.2.0>

* by Philipp Hunziker
* Fast “Raster package” usage

### cellStats::raster

* The raster needs to be cut to the extent of interest, then many statistics can be calculated

### zonal::raster

* [Robert Hijmans](Robert%20Hijmans) in <https://stackoverflow.com/questions/5802649/zonal-statistics-qgis>
* Uses two rasters, i.e. the zones need to be rasterized.

library(raster)

v <- raster('raster filename')

z <- raster('zones raster filename')

zv <- zonal(v, z, fun=mean)

## GRASS

* Mentioned by Paulo in <https://stackoverflow.com/questions/5802649/zonal-statistics-qgis>

# Minimum viable product

* Image Name convention:
  + a190518esccel.tif -> [1 cam][6 yymmdd date][3 trial][3 trait].[ext]
* Vectors Name convention:
  + esc190518bufa.shp -> [3 trial][6 yymmdd date] [3 trait][1 cam\*].[ext]
* Extract images, give the list of images paths and corresponding zone vectors.
  + Allow multiband

# Thanks to:

## https://stackoverflow.com/questions/11007178/creating-a-prompt-answer-system-to-input-data-into-r

## For the idea of the menu.