**MRes mini project: Accessibility stages document**

*Create a document that lists each of the stages involved in generating and summarising the accessibility surface.*

**Need to fill in ‘packages’ columns**

**Need to add info as to what is being done**

**Need to complete ‘setup’ section**

**Re-number all steps**

**1. Setup**

As part of the MRes mini-project, the .Rmd code was restructured to contain an initial ‘setup’ section, allowing users to input specific parameters relating to:

* Downloading/installing/loading all packages needed to generate and summarise the accessibility surface, including those potentially missing from client computers, from both CRAN and GitHub servers (where not available from CRAN)
* Connecting to and initialising rgee package
* Defining the area of interest (through inputting specific co-ordinate data or shapefiles)
* Reading in Landsat 8 data (acquired using Google Earth Engine (GEE))
* NDVI
* Walking travel speeds (non-vehicle, km/ hour) expected when traversing non-road pixels (dependant on NDVI value and customisable according to expected changes in travel speed, e.g., during the wet season.
* Road travel speeds (by motor vehicle, km/ hour) expected when traversing road pixels (obtained using [www.openstreetmaps.org](http://www.openstreetmaps.org) and customisable according to expected changes in travel speed, e.g., during the wet season).
* Population data, obtained using the wpgpDownloadR package, an interface for downloading raster population data from the WorldPop FTP.
* Reading in health facility data (afrimapr?)

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Download/install/load all required packages available from CRAN. | **Packages to install:**  sf,  mapview,  googledrive,  osmdata,  ggplot2,  raster,  gdistance,  fasterize,  remotes,  rgdal,  stars,  geojsonio  devtools  rgee | # Create vector list of packages needed  list.of.packages <-  c(list\_packages)  # Identify which packages aren’t currently installed on client computers and store in object  new.packages <-  list.of.packages[!(list.of.packages  %in% installed.packages()[  ,”Package”]))  # Install any missing packages  if(length(new.packages))  install.packages(new.packages)  # Load all required packages  lapply(list.of.packages, library,  character.only = TRUE) | Executed within R,  Does **not** require GEE |
| Download, install and load wpgpDownloadR package (not yet available from CRAN and so is downloaded from GitHub using devtools package). | **Packages to install:**  wpgpDownloadR | # Download and install wpgpDownloadR package  Devtools::install\_github(  “wpgp/wpgpDownloadR)  # Load wpgpDownloadR package  library(wpgpDownloadR) | Executed within R,  Does **not** require GEE |
| Connect to and initialize rgee package.  More info on this process can be found here: | rgee  googledrive | # Connect to Google Earth Engine (GEE) using rgee (only need to run once)  ee\_install(rgee)  # Initialize rgee package to check whether everything has been set up correctly in order to begin using Google Earth Engine (GEE) via R (only need to setup once)  ee\_initialize(drive = TRUE) | Executed within R,  **Performed using GEE** |

**4. Define area of interest (aoi) and create aoi polygon**

Area of interest defined using ‘WGS84’ projection.

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Define area of interest (aoi) using coordinates (WGS84 projection)  Maximum and minimum values for LAT/LONG; four coordinates used to make square: (bbxmin, bbxmax, bbymin, bbymax) |  | bbxmin <- co-ordinate\_1\_xmin (WGS84)  bbxmax <- co-ordinate\_2\_xmax (WGS84)  bbymin <- co-ordinate\_3\_ymin (WGS84)  bbymax <- co-ordinate\_4\_ymax (WGS84) | Executed within R,  Does **not** require GEE |
| Transform aoi into polygon | rgee | aoi <-  ee$Geometry$Polygon(  cords = list (c(bbxmin, bbymax),  c(bbxmax, bbymax),  c(bbxmax, bbymin),  c(bbxmin, bbymin))) | Executed within R,  **Performed using GEE** |

**5. Landsat 8 data: Read in Landsat 8 (LS8) Tier 1 dataset and filter by area of interest and collection date**

Acquiring Landsat 8 data requires use of GEE?

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Read in Landsat 8 (LS8) Tier 1 dataset | rgee | ls8 <-  ee$ImageCollection(landsat\_data) | Executed within R,  **Performed using GEE** |
| Filter LS8 data by aoi |  | spatialFiltered <-  ls8$filterBounds(aoi) | Executed within R,  **Performed using GEE** |
| Filter LS8 data by collection date |  | temporalFiltered <-  spatialFiltered$filterDate(  date(s)\_of\_interest) | Executed within R,  **Performed using GEE** |

**6. Create and apply a cloud mask to filtered LS8 data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Create cloud mask  (generate function) | rgee | ndvilowcloud <-  function(image) {  # Get a cloud score in [0, 100]  cloud <-  ee$Algorithims$landset$simpleCloudScore(  image)$select(‘cloud’)  # Create a mask of cloudy pixels from an arbitrary threshold (20%)  mask <- cloud$lte(20)  # Compute NDVI using inbuilt functions  nvdi <-  image$normalizedDifference(  c(‘B5’, ‘B4’))$rename(‘NDVI’)  # Return the masked image with an NDVI band  image$addBands(ndvi)$updateMask(mask)  } | Executed within R,  **Performed using GEE** |
| Apply cloud mask |  | cloudlessNDVI =  temporalFiltered$map(ndvilowcloud) | Executed within R,  Does **not** require GEE |

**7. Normalised difference vegetation index (NDVI): Calculate median NDVI per pixel and clip to area of interest (and check output)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Calculate median NDVI per pixel |  | medianimage <-  cloudlessNDVI$median()$select(‘NDVI’) | Executed within R,  **Performed using GEE** |
| Clip to aoi |  | medNDVIaoi <-  medianimage$clip(aoi) | Executed within R,  **Performed using GEE** |
| View output |  | Map$centerObject(aoi)  Map$addLayer(  eeObject = medNDVIaoi,  viaParam = list(min = -1,  max = 1,  palette = c(‘blue’,  ‘white’,  ‘green’)),  name = “Median NDVI”) | Executed within R,  **Performed using GEE** |

Should produce map of aoi

**8. Convert image to raster and download it using Google Drive (drive) or Google Cloud Storage (GCS)**

These data are saved as an image within GEE.

Convert data to raster and download using drive or GCS.

Raster is stored as .tif file in a temporary local folder, which can then be written to our data folder.

More information on this process can be found here: <https://r-spatial.github.io/rgee/reference/ee_as_raster.html>

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Convert data to raster (within GEE) and download/store in temporary folder |  | med\_ndvi <-  ee\_as\_raster(  image = medNDVIaoi,  region = aoi,  scale = 30  via = ‘drive’) | Executed within R,  **Performed using GEE** |
| Write raster (.tif) to local folder |  | writeRaster(  med\_ndvi,  “local\_filepath”,  Format = ‘GTiff”,  Overwrite = TRUE) | Executed within R,  **Performed using GEE** |

**9. OpenStreetMap (OSM) data: Download OSM road network data for our aoi within R.**

To detail travel speeds within the area of interest, an open source road network which is publicly compiled and hosted by OpenStreetMaps (OSM) is used.

Hosted here: [www.openstreetmap.org](http://www.openstreetmap.org)

OSM road data from the aoi can be directly downloaded within R.

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Define bounding box |  | aoi\_bbox <-  c(bbxmin,  bbymin,  bbxmax,  bbymax) | Executed within R,  Does **not** require GEE |
| Obtain road data | osmdata | q <-  opq(bbox = aoi\_bbox) %>%  add\_osm\_feature(key = ‘highway’) %>%  osmdata\_sf() | Executed within R,  Does **not** require GEE |
| Plot road data to check | ggplot  osmdata | ggplot(q$osm\_lines) + geom\_sf() |  |

**Should produce map of road data**

**10a. Assign speeds to NDVI pixels by walking speed**

Need to document how these speeds are determined and how they vary by season and location.

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Temporarily read in NDVIexample data from folder (downloaded from GEE) and save as raster object  This may be replaced with med\_ndvi if rgee continues to be reliable. |  | ndvipath <- “NDVIexample.tif”  ndvi <- raster(ndvipath) | Executed within R,  Does **not** require GEE |
| Reclassify raster so that:  **< 0.35** = impassable (cannot traverse)  **0.35 – 0.6** = 3.5 km p/hour  **0.6 – 0.7** = 2.48 km p/hour  **> 0.7** = 1.49 km p/hour |  | # Generate ndvi\_walk\_kph vector object  ndviwalk\_kph <- c(0.1,  3.5,  2.48,  1.49)  # Convert ndvi\_walk\_kph vector to metres p/second  ndviwalk\_mps <- ndviwalk\_kph/3.6  # Convert to crossing time in seconds (assumes travel along hypotenuse and pixel size is 30 m2)  nvdiwalk\_secs <- 42.43 / ndviwalk\_mps  ## Convert km p/hour to metres p/second using matrix  # Create matrix  ndviwalk\_vec <-  c(-1, 0.35, nvdiwalk\_secs[1],  0.35, 0.6, nvdiwalk\_secs[2],  0.6, 0.7, nvdiwalk\_secs[3],  0.7, 1, nvdiwalk\_secs[4])  ndviwalk\_mat <-  matrix(ndviwalk\_vec,  ncol = 3,  byrow = TRUE)  # Reclassify ndvi raster  ndvi\_assigned <- ndvi  ndvi\_assigned <-  reclassify(ndvi\_assigned,  ndviwalk\_mat) | Executed within R,  Does **not** require GEE |

**10b. Assign speeds by motor vehicle to OSM road data**

Need to document how these speeds are determined and how they vary by season and location.

For primary and major roads, national speed limits are used as a maximum.

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Create dataframe (to be converted to raster) |  | # Create road-type vector  road\_vector <- c(“primary”,  “secondary”,  “motorway”,  “trunk”)  # Set road speeds to those within road\_vector to 80 km/h, and all other road types to 20 km/h  q$osm\_lines$motorspeedkph <-  ifelse(q$osm\_lines$highway %in%  road\_vector, 80, 20)  # Convert to metres p/second  q$osm\_lines$motorspeedmps <-  q$osm\_lines$motorspeedkph / 3.6  # Assume a 30 m resolution cell  q$osm\_lines$time\_secs <-  42.43 / q$osm\_lines$motorspeedmps | Executed within R,  Does **not** require GEE |
| Convert to raster, matching the NDVI raster resolution and extent | fasterise | # fasterise function only works with polygons, so a road buffer of ~30 m is added  roads.poly <-  st\_buffer(q$osm\_line, 0.00015)  # Convert to raster  osm\_road\_raster <-  fasterise(roads\_poly,  ndvi\_assigned,  “time\_secs”,  fun = ‘min’) | Executed within R,  Does **not** require GEE |

**11. Merge ‘NVDI’ raster (ndvi\_assigned) and road-data raster (osm\_road\_raster)**

This step will merge data for off-road and on-road travel to create one cohesive friction surface.

In areas where off-road and on-road cells overlap, road values will be retained as these will be associated with the lowest cost (i.e., quickest speed).

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Merge ndvi\_assigned and osm\_road\_raster. Maintain the minimum value, i.e., the quickest cell crossing time. |  | friction\_surface\_motor <-  mosaic(osm\_road\_raster,  ndvi\_assigned,  fun = min,  tolerance = 1) | Executed within R,  Does **not** require GEE |
| Save friction\_surface\_motor raster as .tif file |  | writeraster(friction\_surface\_motor,  “local\_filepath”,  format = “GTiff”,  overwrite = TRUE) | Executed within R,  Does **not** require GEE |

**11. Read in health facility location data**

Will be revised in future to use AfriMapR?

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Read in health facility data and store as object |  | healthfac <-  st\_read(“health\_facility\_data.shp”) | Executed within R,  Does **not** require GEE |

**12. Calculate shortest paths**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Calculate transition matrix |  | trans\_motor <-  transition(friction\_surface\_motor,  transitionFunction =  function(x){1/mean(x)},  directions = 8) | Executed within R,  Does **not** require GEE |
| Calculate cumulative cost |  | leastcost\_motor <-  accost(trans\_motor,  as\_Spatial(healthfac)) | Executed within R,  Does **not** require GEE |
| Save leastcost\_motor raster as .tif file |  | writeRaster(leastcost\_motor,  “local\_filepath”,  format = “GTiff”,  overwrite = TRUE) | Executed within R,  Does **not** require GEE |

**13. Create plot to visualise leastcost\_motor raster data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Store leastcost\_motor as dataframe |  | lcm\_df <-  as.data.frame(leastocost\_motor,  xy = TRUE) | Executed within R,  Does **not** require GEE |
| Add ‘mins’ column by dividing ‘Layer’ column (shortest path time in seconds) by 60 |  | Lcm\_df$mins <-  lcm\_df$Layer / 60 | Executed within R,  Does **not** require GEE |
| Create plot to visualise raster data | ggplot | ggplot()+  geom\_raster(  data = lcm\_df,  aes(x = x,  y = y,  fill = cut(mins,  c(0,30,  60, 120  180, 240,  300,  max(mins)))))+  scale\_fill\_brewer(  palette = “YlGnBu”)+  geom\_sf(  data = q$osm\_lines,  colour = “darkgrey”,  alpha = 0.3)+  geom\_sf(  data = healthfac,  size = 2,  colour = “red”)+  guides(fill= guide\_legend(  title= “Time (mins)”)) | Executed within R,  Does **not** require GEE |

**15. Obtain ISO3 (country code) for country of interest**

Using the wpgpListCountries function within the wpgpDownloadR package, download a ISO3 (country code) dataframe from the WorldPop FTP to obtain ISO3 code for country of interest (Malawi).

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Save ISO3 dataframe as object by calling wpgpListCountries function | wpgpDownloadR | ISO3\_df <- wpgpListCountries() | Executed within R,  Does **not** require GEE |
| Ascertain ISO3 country code for country of interest (Malawi\*) and store ISO3 country code as object  \* Country of interest (COI) specified and stored as object during initial setup stages (stage X) | wpgpDownloadR | ISO3 <-  ISO3\_df[ISO3\_df$Country == COI,  “ISO3”] | Executed within R,  Does **not** require GEE |

**16. Download population data (.tif) for country of interest and covariate of interest, and create raster**

Using the wpgpListCountryDatasets function within the wpgpDownloadR package, download population dataset (.tif) for country and covariate of interest (Malawi/total population per grid cell in 2020) form the WorldPop FTP and create raster from downloaded data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Download dataset of available covariates for country of interest (Malawi) and store as object | wpgpDownloadR | covariates\_df <-  wpgpListCountryDatasets(ISO3 = ISO3) | Executed within R,  Does **not** require GEE |
| Check covariates dataframe to identify covariate of interest (total population per grid cell in 2020) and store as object | wpgpDownloadR | # Check covariates dataframe  View(covariates\_df)  # Identify covariate of interest and store as object  covariate <- “covariate\_of\_interest” | Executed within R,  Does **not** require GEE |
| Download dataset (.tif) for country and covariate of interest, using ISO3 and covariate\_of\_interest (downloaded .tif stored locally) | wpgpDownloadR | wpgpGetCountryDataset(  ISO3 = ISO3,  covariate = covariate  destDIR = (“local\_filepath”) | Executed within R,  Does **not** require GEE |
| Read in downloaded .tif file |  | pop\_data <- “local\_filepath” | Executed within R,  Does **not** require GEE |
| Create raster from .tif file |  | pop\_data <- raster(pop\_data) | Executed within R,  Does **not** require GEE |

**17. Resample and clip pop\_data raster to match resolution and extent of leastcost\_motor raster**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Check pop\_data resolution |  | res(pop\_data)  # 0.0008333333 0.0008333333 | Executed within R,  Does **not** require GEE |
| Check leastcost\_motor resolution |  | res(leastcost\_motor)  # 0.0002694946 0.0002694946 | Executed within R,  Does **not** require GEE |
| Use resample function to: \* Resample pop\_data to match resolution of leastcost\_motor  \* Clip pop\_data to match extent of leastcost\_motor |  | pop\_data <-  resample(pop\_data,  leastcost\_motor,  method = “bilinear”) | Executed within R,  Does **not** require GEE |
| Check resolution and extent of pop\_data match that of leastcost\_motor |  | # Check resolution  res(pop\_data) ==  res(leastcost\_motor)  # TRUE  # Check extent  extent(pop\_data) == extent(leastcost\_motor)  # TRUE | Executed within R,  Does **not** require GEE |

**18. Reclassify leastcost\_motor raster to chosen time-boundary categorical zones**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Create raster from lcm\_df |  | # rasterFromXYZ function works only with three columns, so remove ‘layer’ column from lcm\_df  lcm\_df$layer <- NULL  # Create raster  lcm\_raster <- rasterfromXYZ(lcm\_df) | Executed within R,  Does **not** require GEE |
| Create matrix of time-boundary categories of interest (This will be used to resample lcm\_raster) |  | ## Categories:  # < 30 minutes  # 30 minutes – 1 hour  # 1 hour – 3 hours  # 3 hours – 6 hours  # 6 hours – 12 hours  # > 12 hours  rcl\_matrix <- c( 0, 30, 1,  30, 60, 2,  60, 180, 3,  180, 360, 4,  360, 720, 5,  720, 1000000, 6)  # Reorder rcl\_matrix  rcl\_matrix <- matrix(rcl\_matrix,  ncol = 3,  byrow = TRUE) | Executed within R,  Does **not** require GEE |
| Reclassify lcm\_raster using lcm\_matrix according to time-boundary categorical zones |  | lcm\_pop\_data\_rcl <-  reclassify(lcm\_raster,  rcl\_matrix,  include.lowest = TRUE) | Executed within R,  Does **not** require GEE |
| Assign lcm\_pop\_data\_rcl coordinate reference system (CRS) to that of leastcost\_motor (CRS: WGS84) |  | # Assign CRS  projection(lcm\_pop\_data\_rcl) <-  crs(leastcost\_motor)  # Check CRS  crs(lcm\_pop\_data\_rcl)  # WGS84 | Executed within R,  Does **not** require GEE |
| Plot lcm\_pop\_data\_rcl to visualise |  | plot(lcm\_pop\_data\_rcl) | Executed within R,  Does **not** require GEE |

**19. Determine population within each time-boundary zone and summate data.**

Summating population data within each time-boundary zone allows for number of people within certain travel times from health facilities to be determined and also allows for the percent of the population within certain travel times from health facilities to be calculated.

Create for-loops?

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Package(s)** | **Function(s)** | **R/GEE** |
| Determine population (population data within pop\_data raster) within time-boundary zones using lcm\_pop\_data\_rcl raster and zonal function. |  | lcm\_rcl\_zone <-  zonal(pop\_data,  lcm\_pop\_data\_rcl,  fun = sum) | Executed within R,  Does **not** require GEE |
| Create dataframe from lcm\_rcl\_zone |  | # Create dataframe  lcm\_rcl\_zone\_df <-  as.data.frame(lcm\_rcl\_zone)  # Rename columns  names(lcm\_rcl\_zone\_df)[1] <-  “Zone”  names(lcm\_rcl\_zone\_df)[2] <-  “Zone Population”  # Replace time-boundary zone codes with chosen time-boundary categories  lcm\_rcl\_zone\_df$Zone <-  c(“< 30 minutes,  30 minutes – 1 hour,  1 hour – 3 hours,  3 hours – 6 hours,  6 hours – 12 hours,  > 12 hours)  # Add ‘Total Population’ column  lcm\_rcl\_zone\_df$”Total Population” <-  c(sum(  lcm\_rcl\_zone\_df$”Zone Population”[1]),  lcm\_rcl\_zone\_df$”Zone Population”[1:2]),  lcm\_rcl\_zone\_df$”Zone Population”[1:3]),  lcm\_rcl\_zone\_df$”Zone Population”[1:4]),  lcm\_rcl\_zone\_df$”Zone Population”[1:5]),  lcm\_rcl\_zone\_df$”Zone Population”[1:6]))  # Add ‘% of Total Population’ column  lcm\_rcl\_zone\_df$”% Population” <-  c(sum(  lcm\_rcl\_zone\_df$”Zone Population”[1]/  lcm\_rcl\_zone\_df$”Zone Population”[1:6])  \* 100,  lcm\_rcl\_zone\_df$”Zone Population”[1:2]/  lcm\_rcl\_zone\_df$”Zone Population”[1:6])  \* 100,  lcm\_rcl\_zone\_df$”Zone Population”[1:3]/  lcm\_rcl\_zone\_df$”Zone Population”[1:6])  \* 100,  lcm\_rcl\_zone\_df$”Zone Population”[1:4]/  lcm\_rcl\_zone\_df$”Zone Population”[1:6])  \* 100,  lcm\_rcl\_zone\_df$”Zone Population”[1:5]/  lcm\_rcl\_zone\_df$”Zone Population”[1:6])  \* 100,  lcm\_rcl\_zone\_df$”Zone Population”[1:6]/  lcm\_rcl\_zone\_df$”Zone Population”[1:6])  \* 100) | Executed within R,  Does **not** require GEE |
| View lcm\_rcl\_zone\_df dataframe |  | View(lcm\_rcl\_zone\_df) |  |