

# GLUE-urbanQuant

Data Management

David Murray-Stoker

## Contents

<b>Load Data</b>	<b>2</b>
GLUE Transects . . . . .	3
Random Transects . . . . .	5
Random Points . . . . .	7
Systematic Points . . . . .	9
<b>Manage Data for Analyses</b>	<b>11</b>
<b>Workspace Information</b>	<b>12</b>

## Load Data

```
## Creates a list with dataframes as elements for all csv files in data_in
##
## @param data_in Path to directory containing csv files to load
##
## @return df_list List of dataframes, each from a different csv file

## Set the function
create_df_list <- function(data_in) {
  # Get all csv files in data_in
  files <- dir(data_in, pattern = "*.csv")

  # Read in all the files, appending the path before the filename
  df_list <- files %>%
    map(~ read_csv(file.path(data_in, .), show_col_types = FALSE))

  return(df_list)
}
```

## GLUE Transects

```
## Set paths for the GLUE transects
GLUE.transects.Distance.path <- "data/distance_data/GLUE_transects_Distance/"
GLUE.transects.ISC.path <- "data/ISC_data/GLUE_transects_ISC/"
GLUE.transects.HII.path <- "data/HII_data/GLUE_transects_HII"
GLUE.transects.NDVI.path <- "data/NDVI_data/GLUE_transects_NDVI/"
GLUE.transects.WorldClim.path <- "data/WorldClim_data/GLUE_transects_WorldClim/"
GLUE.transects.Aridity_Index.path <- "data/Aridity_Index_data/GLUE_transects_Aridity_Index/"
GLUE.transects.GDP.SSP.path <- "data/GDP_SSP_data/GLUE_transects_GDP_SSP"

## Import data for the GLUE transects
# Distance
GLUE.transects.distance.data <- create_df_list(GLUE.transects.Distance.path) %>%
  bind_rows()
# ISC
GLUE.transects.ISC.data <- create_df_list(GLUE.transects.ISC.path) %>%
  bind_rows() %>%
  select(UID, ISC_Mean)
# HII
GLUE.transects.HII.data <- create_df_list(GLUE.transects.HII.path) %>%
  bind_rows() %>%
  select(UID, HII)
# NDVI
GLUE.transects.NDVI.data <- create_df_list(GLUE.transects.NDVI.path) %>%
  bind_rows() %>%
  select(UID, Mean_NDVI, Min_NDVI, Max_NDVI)
# WorldClim
GLUE.transects.WorldClim.data <- create_df_list(GLUE.transects.WorldClim.path) %>%
  bind_rows() %>%
  select(-c(Country, City))
# Aridity Index
GLUE.transects.Aridity_Index.data <- create_df_list(GLUE.transects.Aridity_Index.path) %>%
  bind_rows() %>%
  select(UID, Aridity_Index)
# GDP & SSP
GLUE.transects.GDP.SSP.data <- create_df_list(GLUE.transects.GDP.SSP.path) %>%
  bind_rows() %>%
  select(UID, GDP_2005:SSP_5_2100)

## Merge and organize full data
GLUE.transects.full.data <- GLUE.transects.distance.data %>%
  full_join(GLUE.transects.ISC.data, by = "UID") %>%
  full_join(GLUE.transects.HII.data, by = "UID") %>%
  full_join(GLUE.transects.NDVI.data, by = "UID") %>%
  full_join(GLUE.transects.WorldClim.data, by = "UID") %>%
  full_join(GLUE.transects.Aridity_Index.data, by = "UID") %>%
  full_join(GLUE.transects.GDP.SSP.data, by = "UID") %>%
  select(
    UID, City, Distance, ISC_Mean, HII,
    Mean_NDVI, Min_NDVI, Max_NDVI,
    Mean_Annual_Temperature, Temperature_Seasonality, Range_Annual_Temperature,
    Annual_Precipitation, Precipitation_Seasonality, Aridity_Index, GDP_2005:SSP_5_2100
  )
```

```
## Clarify UID for later analyses  
GLUE.transects.full.data$UID <- paste(GLUE.transects.full.data$UID, "GLUE", sep = ".")
```

## Random Transects

```
## Set paths for the random transects data
random.transects.Distance.path <- "data/Distance_data/random_transects_Distance/"
random.transects.ISC.path <- "data/ISC_data/random_transects_ISC/"
random.transects.HII.path <- "data/HII_data/random_transects_HII/"
random.transects.NDVI.path <- "data/NDVI_data/random_transects_NDVI/"
random.transects.WorldClim.path <- "data/WorldClim_data/random_transects_WorldClim/"
random.transects.Aridity_Index.path <- "data/Aridity_Index_data/random_transects_Aridity_Index/"
random.transects.GDP.SSP.path <- "data/GDP_SSP_data/random_transects_GDP_SSP"

## Import data for the random transects
# Distance
random.transects.distance.data <- create_df_list(random.transects.Distance.path) %>%
  bind_rows()
# ISC
random.transects.ISC.data <- create_df_list(random.transects.ISC.path) %>%
  bind_rows() %>%
  select(UID, ISC_Mean)
# HII
random.transects.HII.data <- create_df_list(random.transects.HII.path) %>%
  bind_rows() %>%
  select(UID, HII)
# NDVI
random.transects.NDVI.data <- create_df_list(random.transects.NDVI.path) %>%
  bind_rows() %>%
  select(UID, Mean_NDVI, Min_NDVI, Max_NDVI)
# WorldClim
random.transects.WorldClim.data <- create_df_list(random.transects.WorldClim.path) %>%
  bind_rows() %>%
  select(-c(Country, City))
# Aridity Index
random.transects.Aridity_Index.data <- create_df_list(random.transects.Aridity_Index.path) %>%
  bind_rows() %>%
  select(UID, Aridity_Index)
# GDP & SSP
random.transects.GDP.SSP.data <- create_df_list(random.transects.GDP.SSP.path) %>%
  bind_rows() %>%
  select(UID, GDP_2005:SSP_5_2100)

## Merge and organize full random.transects data
random.transects.full.data <- random.transects.distance.data %>%
  full_join(random.transects.ISC.data, by = "UID") %>%
  full_join(random.transects.HII.data, by = "UID") %>%
  full_join(random.transects.NDVI.data, by = "UID") %>%
  full_join(random.transects.WorldClim.data, by = "UID") %>%
  full_join(random.transects.Aridity_Index.data, by = "UID") %>%
  full_join(random.transects.GDP.SSP.data, by = "UID") %>%
  select(
    UID, City, Distance, ISC_Mean, HII,
    Mean_NDVI, Min_NDVI, Max_NDVI,
    Mean_Annual_Temperature, Temperature_Seasonality, Range_Annual_Temperature,
    Annual_Precipitation, Precipitation_Seasonality, Aridity_Index, GDP_2005:SSP_5_2100
  )
```

```
## Clarify UID for later analyses  
random.transects.full.data$UID <- paste(random.transects.full.data$UID, "RT", sep = ".")
```

## Random Points

```
## Set paths for the random points data
random.points.Distance.path <- "data/Distance_data/random_points_Distance/"
random.points.ISC.path <- "data/ISC_data/random_points_ISC/"
random.points.HII.path <- "data/HII_data/random_points_HII/"
random.points.NDVI.path <- "data/NDVI_data/random_points_NDVI/"
random.points.WorldClim.path <- "data/WorldClim_data/random_points_WorldClim/"
random.points.Aridity_Index.path <- "data/Aridity_Index_data/random_points_Aridity_Index/"
random.points.GDP.SSP.path <- "data/GDP_SSP_data/random_points_GDP_SSP"

## Import data for the random points
# Distance
random.points.distance.data <- create_df_list(random.points.Distance.path) %>%
  bind_rows()
# ISC
random.points.ISC.data <- create_df_list(random.points.ISC.path) %>%
  bind_rows() %>%
  select(UID, ISC_Mean)
# HII
random.points.HII.data <- create_df_list(random.points.HII.path) %>%
  bind_rows() %>%
  select(UID, HII)
# NDVI
random.points.NDVI.data <- create_df_list(random.points.NDVI.path) %>%
  bind_rows() %>%
  select(UID, Mean_NDVI, Min_NDVI, Max_NDVI)
# WorldClim
random.points.WorldClim.data <- create_df_list(random.points.WorldClim.path) %>%
  bind_rows() %>%
  select(-c(Country, City))
# Aridity Index
random.points.Aridity.Index.data <- create_df_list(random.points.Aridity_Index.path) %>%
  bind_rows() %>%
  select(UID, Aridity_Index)
# GDP & SSP
random.points.GDP.SSP.data <- create_df_list(random.points.GDP.SSP.path) %>%
  bind_rows() %>%
  select(UID, GDP_2005:SSP_5_2100)

## Merge and organize full random.points data
random.points.full.data <- random.points.distance.data %>%
  full_join(random.points.ISC.data, by = "UID") %>%
  full_join(random.points.HII.data, by = "UID") %>%
  full_join(random.points.NDVI.data, by = "UID") %>%
  full_join(random.points.WorldClim.data, by = "UID") %>%
  full_join(random.points.Aridity.Index.data, by = "UID") %>%
  full_join(random.points.GDP.SSP.data, by = "UID") %>%
  select(
    UID, City, Distance, ISC_Mean, HII,
    Mean_NDVI, Min_NDVI, Max_NDVI,
    Mean_Annual_Temperature, Temperature_Seasonality, Range_Annual_Temperature,
    Annual_Precipitation, Precipitation_Seasonality, Aridity_Index, GDP_2005:SSP_5_2100
  )
```

```
## Clarify UID for later analyses  
random.points.full.data$UID <- paste(random.points.full.data$UID, "RP", sep = ".")
```



## Systematic Points

```
## Set paths for the systematic points data
systematic.points.Distance.path <- "data/Distance_data/systematic_points_Distance/"
systematic.points.ISC.path <- "data/ISC_data/systematic_points_ISC/"
systematic.points.HII.path <- "data/HII_data/systematic_points_HII/"
systematic.points.NDVI.path <- "data/NDVI_data/systematic_points_NDVI/"
systematic.points.WorldClim.path <- "data/WorldClim_data/systematic_points_WorldClim/"
systematic.points.Aridity_Index.path <- "data/Aridity_Index_data/systematic_points_Aridity_Index/"
systematic.points.GDP.SSP.path <- "data/GDP_SSP_data/systematic_points_GDP_SSP"

## Import data for the systematic points
# Distance
systematic.points.distance.data <- create_df_list(systematic.points.Distance.path) %>%
  bind_rows()
# ISC
systematic.points.ISC.data <- create_df_list(systematic.points.ISC.path) %>%
  bind_rows() %>%
  select(UID, ISC_Mean)
# HII
systematic.points.HII.data <- create_df_list(systematic.points.HII.path) %>%
  bind_rows() %>%
  select(UID, HII)
# NDVI
systematic.points.NDVI.data <- create_df_list(systematic.points.NDVI.path) %>%
  bind_rows() %>%
  select(UID, Mean_NDVI, Min_NDVI, Max_NDVI)
# WorldClim
systematic.points.WorldClim.data <- create_df_list(systematic.points.WorldClim.path) %>%
  bind_rows() %>%
  select(-c(Country, City))
# Aridity Index
systematic.points.Aridity.Index.data <- create_df_list(systematic.points.Aridity_Index.path) %>%
  bind_rows() %>%
  select(UID, Aridity_Index)
# GDP & SSP
systematic.points.GDP.SSP.data <- create_df_list(systematic.points.GDP.SSP.path) %>%
  bind_rows() %>%
  select(UID, GDP_2005:SSP_5_2100)

## Merge and organize full systematic.points data
systematic.points.full.data <- systematic.points.distance.data %>%
  full_join(systematic.points.ISC.data, by = "UID") %>%
  full_join(systematic.points.HII.data, by = "UID") %>%
  full_join(systematic.points.NDVI.data, by = "UID") %>%
  full_join(systematic.points.WorldClim.data, by = "UID") %>%
  full_join(systematic.points.Aridity.Index.data, by = "UID") %>%
  full_join(systematic.points.GDP.SSP.data, by = "UID") %>%
  select(
    UID, City, Distance, ISC_Mean, HII,
    Mean_NDVI, Min_NDVI, Max_NDVI,
    Mean_Annual_Temperature, Temperature_Seasonality, Range_Annual_Temperature,
    Annual_Precipitation, Precipitation_Seasonality, Aridity_Index, GDP_2005:SSP_5_2100
  )
```

```
## Clarify UID for later analyses  
systematic.points.full.data$UID <- paste(systematic.points.full.data$UID, "SP", sep = ".")
```

## Manage Data for Analyses

```
## Bind all data
full.data.uncleaned <- GLUE.transects.full.data %>%
  bind_rows(random.points.full.data) %>%
  bind_rows(systematic.points.full.data) %>%
  bind_rows(random.transects.full.data)

## Set vector of sample type (GLUE, Random_Points, Systematic_Points, Random_Transect)
full.data.uncleaned$Sampling_Design <- c(
  rep("GLUE", length.out = length(GLUE.transects.full.data$UID)),
  rep("Random_Points", length.out = length(random.points.full.data$UID)),
  rep("Systematic_Points", length.out = length(systematic.points.full.data$UID)),
  rep("Random_Transect", length.out = length(random.transects.full.data$UID))
)

## Standardize distance to scale to a maximum of 1
full.data.uncleaned <- full.data.uncleaned %>%
  group_by(Sampling_Design, City) %>%
  mutate(Standardized_Distance = (Distance / max(Distance)))

## Reorganize the final data
full.data.cleaned <- full.data.uncleaned %>%
  select(
    UID, City, Sampling_Design, Distance, Standardized_Distance,
    ISC_Mean, HII, Mean_NDVI, Min_NDVI, Max_NDVI,
    Mean_Annual_Temperature, Temperature_Seasonality, Range_Annual_Temperature,
    Annual_Precipitation, Precipitation_Seasonality, Aridity_Index,
    GDP_2005:SSP_5_2100
  )
```

## Workspace Information

Table 1: Packages required for data management and analyses.

Package	Loaded Version	Date
dplyr	1.1.4	2023-11-17
forcats	1.0.0	2023-01-29
ggplot2	3.4.4	2023-10-12
kableExtra	1.4.0	2024-01-24
knitr	1.45	2023-10-30
lubridate	1.9.3	2023-09-27
purrr	1.0.2	2023-08-10
readr	2.1.5	2024-01-10
stringr	1.5.1	2023-11-14
tibble	3.2.1	2023-03-20
tidyr	1.3.1	2024-01-24
tidyverse	2.0.0	2023-02-22