# Topological Data Analysis with R workshop

January 2017

Peter Bubenik

## 1. Introduction

In this workshop, you will use R and some persistent homology software to perform topological data analysis. This document is available as pdf with clickable links on the Software page of my webpage `http://people.clas.ufl.edu/peterbubenik/`.

## 2. Getting started

2.1. **R and RStudio.** First, download R from `https://cran.rstudio.com/`, then download RStudio Desktop from `https://www.rstudio.com/products/rstudio/download/`.

If you are new to R, work through the R introduction at `http://www.r-tutor.com/r-introduction`.

2.2. **Folder structure.** Create a directory (folder) for this workshop using your operating system (e.g. windows). For example, `C:\Users\Peter\Desktop\R\workshop` or `~/R/workshop`. Next create a subdirectory named `bin` of this working directory.

2.3. **Persistence software.**

2.3.1. *Perseus.* Download the executable file for your operating system from `http://people.maths.ox.ac.uk/nanda/perseus/` and put in the `bin` folder. For Mac/Linux you will need to make the file executable. For example, `chmod u+x perseusMac`.

2.3.2. *Persistence Landscape Toolbox.* Download the appropriate executables from `https://www.math.upenn.edu/~dlotko/persistenceLandscape.html`, extract the contents and put the files `PlotOfLandscape` and `configure` in the `bin` folder. For Windows, in addition copy the files `libgcc_s_dw2-1.dll` and `libstdc++-6.dll` into the `bin` folder. For Mac/Linux you will need to make the file executable: `chmod u+x PlotOfLandscape`.

2.4. **For Mac users.** You need to install XQuartz from `https://www.xquartz.org/`.

2.5. **The Data.** Download the file `nonrigid3d.zip` from `http://tosca.cs.technion.ac.il/book/resources_data.html`, extract the contents and move the resulting `nonrigid3d` folder to be a subdirectory of your working directory.

2.6. **My R files.** Download the following R files: `tda_workshop_script.R`, `tda_functions.R`, and `persistence_script.R`, from my web page `http://people.clas.ufl.edu/peterbubenik/software/` and put them in your workshop directory. Rename the files from ending in `.txt` to ending in `.R`.

## 3. Doing some Topological Data Analysis with R

**Hints.**

- You can step through the script by executing each line with `<Control>-<Return>`.
- After stepping through a script one command at a time once, you can run the entire script at once using `source("workshop_script.R")`, for example.
- The following commands are particularly useful for resetting RStudio: `rm(list=ls())` and `graphics.off()`.

1. Open the R file `workshop_script.R`. After a summary of what the script does, it contains a number of parameters that the user may change. Change the value of the parameter `main_directory` to the path of your workshop directory. Also change the parameter `operating_system` to the correct value.
2. Step through the commands in `workshop_script.R` one at a time using `<Ctrl><Enter>` or `<Cmnd><Enter>`. Try and understand how each command works and what it is doing. Stop after you execute the line `source("persistence_script.R")`.
3. Once you have executed `persistence_script.R`, you should have 3d views of some of the figures. You should be able to rotate these and zoom in. Now go to the file `workshop_script.R` and change `show_individual_figures <- TRUE` from `TRUE` to `FALSE`. Also increase `max_pose_num` to 50. Make sure you execute these two lines to get the new values into memory. Now rerun `source("persistence_script.R")`.
4. You should now have the persistence landscapes of enough figures to do some statistics and machine learning. Step through the remaining lines in `workshop_script.R`.
5. Now that you have compared two of the figures: change `figures_to_compare` to change the pair of figures to compare; change `figures_to_use` (this can be a list of more than two types) to change the list of figures on which to apply classification; and and rerun your topological data analysis.

If you are interested, take a careful look `persistence_script.R` and at the functions in the file `tda_functions.R` and figure out how they work.

**Bonus.** There is an easter egg in the code that will generate animations of the filtered simplicial complexes of the figures. See if you can get it to run. Note: the saved animated gif isn't cropped properly. See if you can fix this.

## 4. To do

Your assignment consists of doing some of the following.

1. Experiment with comparing and classifying various combinations of the figures.
2. Understand the pipeline used in the analysis.
3. Understand in detail some part of the analysis.
4. Modify the R code to use a different filtering function.
5. Modify the R code to analyze some new synthetic or experimental data.
6. Extend the R code to add some new feature.