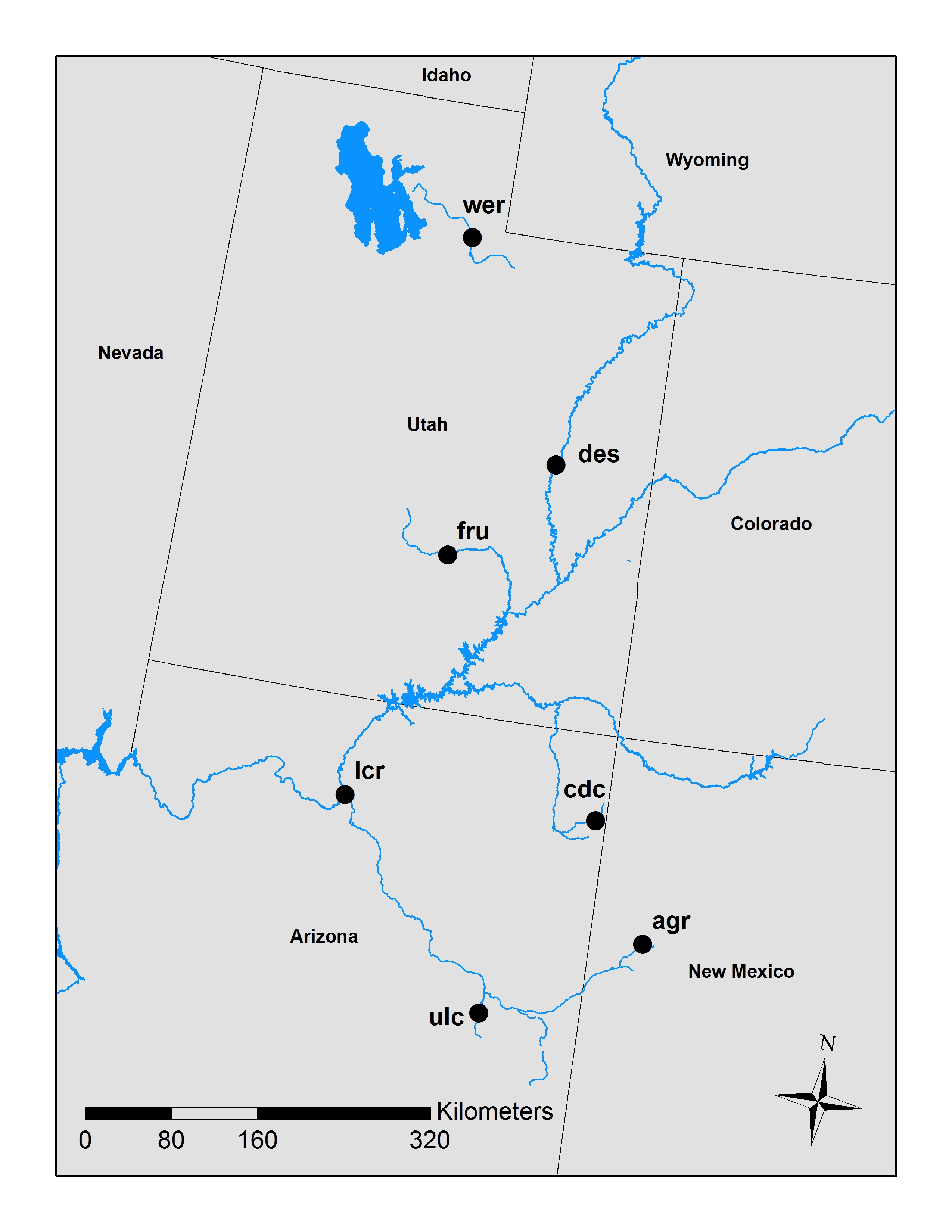
**Computer Lab 9 – Genetic Distance Using Microsatellites**

**Conservation Genetics (BIOL 4174 / 5174)**

**Population Information and Descriptive Statistics**

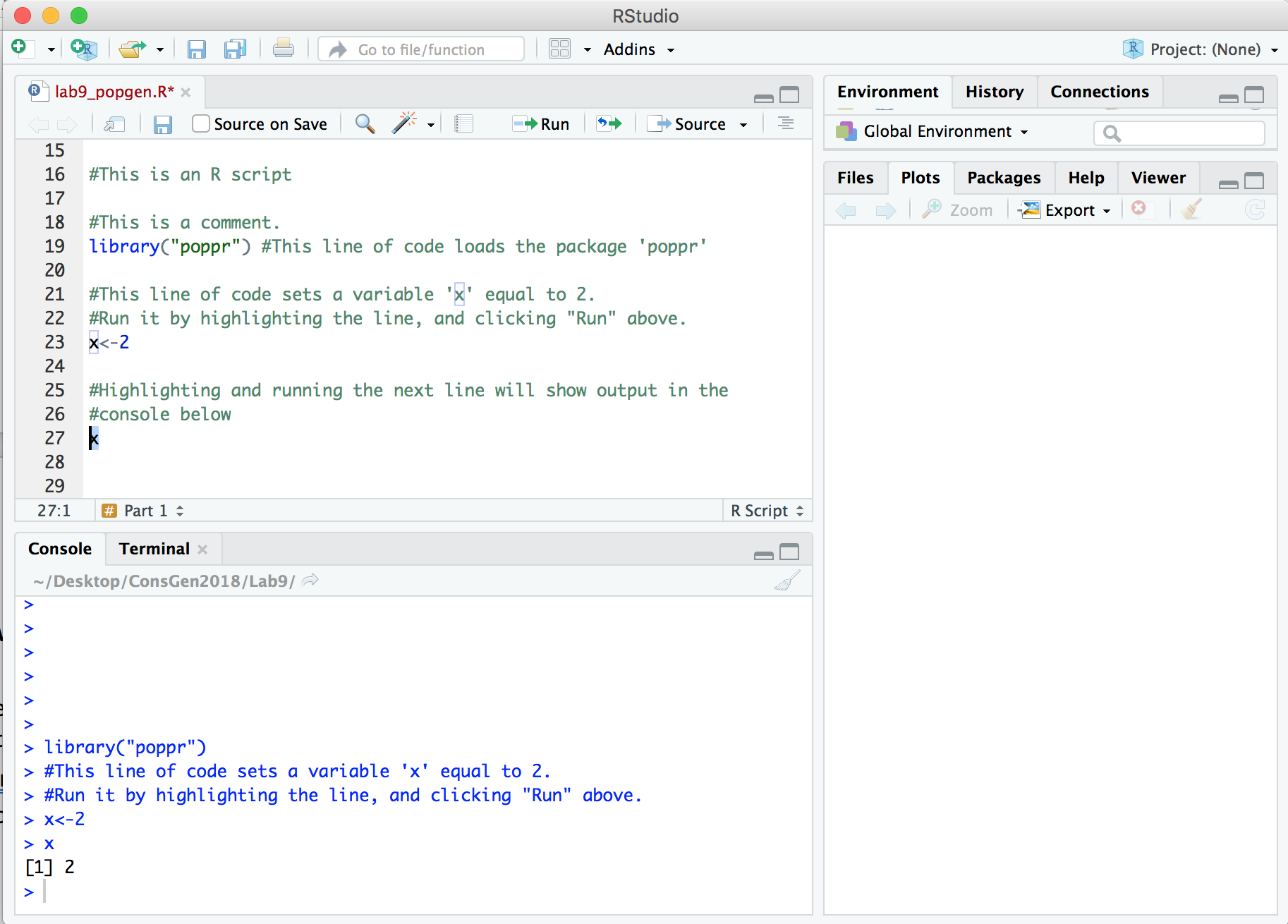
Populations of a native fish (Bluehead Sucker) are distributed throughout the Colorado River (populations 1-6) and Bonneville (population 7) basins in the Southwestern US (see map below). The Bonneville basin has no contemporary connections to other basins or oceans.



1. **AGR** –supposed to be a different subspecies.
2. **CDC** – potentially introgressed by another ESU.
3. **DES** – mainstem population, upper basin.
4. **FRU** –tributary population, upper basin.
5. **LCR** –tributary population, lower basin.
6. **ULC** – different ESU.
7. **WER** – different ESU. Bonneville Basin.

This week we will perform all of our calculations in R. Open R Studio and set your working directory to the location of your Lab 9 files. In your Lab 9 files, you will notice a file “lab9\_popgen.R”. Use the  icon to load this file as an R script. You can then highlight lines of code and copy/paste into your R terminal, or run them using the  icon.

You should be able to complete today’s lab using just the comments (marked with “#”) in the script.



Maximize your R Studio window and follow the instructions in the R script for the remainder of today’s lab. ☺

Reference

* Nei’s Standard Genetic Distance – this measure requires the calculation of **Nei’s Genetic Identity (*I*)**, after which distance is easily calculated using the formula *D*=-ln(*I*). This assumes the markers used are selectively neutral. It is based on the **Infinite Alleles Model** (= IAM): all mutations result in new alleles of any size; thus mutations can consist of addition or deletion of any numbers of repeats. Genetic Identity ranges from 0 to 1, while Genetic Distance ranges from 0 to infinity. Negative values are interpreted as 0.
* *G*ST – this measure is the proportion of genetic diversity found within populations. When only two alleles exist at a locus *G*ST = *F*ST. When there are multiple at a locus, *G*ST is the weighted average of *F*ST for all alleles.
* *F*ST – this is a measure of the variance among subpopulations relative to the total population variance. *F*ST assumes an infinite alleles model (IAM).
* *G*ST’ (prime) – A ‘standardizes’ Gst calculated by dividing Gst for each marker by the maximum theoretical Gst based on the heterozygosity of that marker.
* *R*ST – analog of *F*ST adapted specifically for microsatellite loci, except it assumes a **stepwise mutation model** (= SMM); mutations consist of addition or deletion of a single repeat. Consequently, alleles of similar size are genetically more closely “related” than alleles of very different size.
* HWE – A population in Hardy-Weinberg equilibrium have allele and genotype frequencies conforming to the mathematical expectation under an idealized population- one with random mating, a large population size, no mutation, no selection, no population structure, and no gene flow.
* Linkage Disequilibrium– Non-random association of alleles. Can be caused by physical linkage, co-adaptation, or as an artefact of sampling.
* – The most commonly applied index of association, used to test for LD. It is computed using the covariance in genetic distance among loci