Appendix B.

Development of Hydrogeographic Fish Assemblages for California

### Nick Santos, Ryan Peek, Rob Lusardi and Sarah Yarnell

### Center for Watershed Sciences, UC Davis

# Introduction

The identification of priority ecological endpoints across the state is a key step in refining Tier 1 environmental flow recommendations as outlined in Tier 2 of the Calfornia Environmental Flows Framework (CEFF) under development by the E-flows Technical Workgroup, a subgroup of the California Water Quality Monitoring Council. Based on a literature review of documented relationships between aquatic species and flow metrics and conditions, we found that data directly linking individual species to quantifiable flow metrics is available for a few key species of concern, but otherwise is broadly lacking (Yarnell et al., submitted). Therefore, an approach is needed to more broadly relate stream flow conditions to aquatic species communities and the variety of species life history needs within a given community.

Here, we begin to address this need by developing an approach for determining regional fish assemblages across California that are geographically distinct and thus are assumed to have co-evolved under similar climate conditions. We use fish distribution data from the PISCES database (Santos et al, 2014) and cluster species geographically within broad regions across California. Because PISCES includes all fish species within the state and provides designations of native, non-native, historic, current, and threatened and endangered (T&E) status, we sought to first narrow the number of species assessed to only those species known to most closely associate with certain flow conditions. These “flow sensitive species” are thus most susceptible to altered flow regimes and can most benefit from improved environmental flow regimes (Grantham et al, 2014).

A variety of methods are available for clustering data, and we chose to use a simple k-means clustering approach to determine fish assemblages. Because we ultimately would like to compare the fish assemblages with hydrologic conditions, such as stream class and unimpaired flow metrics, we wanted the clustering to be based on geographic proximity of species alone. This approach allows us to determine clusters of species that are most distinct within a defined region, based on their inherent geographic similarity and their dissimilarity with other clusters within the region. The k-means clustering approach is highly dependent on the number of clusters manually selected for analysis, thus we evaluated a range of cluster sets (number of clusters = 2-9) for each region. In order to determine which set of clusters within a region was most geographically sound and ecologically relevant, we assessed the various clusters visually and via ordination plots, then used expert opinion to select a final set of clusters for each region. Because we focused on flow sensitive fish species and their geographic proximity within a region to define clusters, we term each resulting cluster in a region as a “hydrogeographic fish assemblage”.

Details on the methods used to determine the hydrogeographic fish assemblages is provided below. The results, including an example of the analysis completed to inform cluster selection, is also provided below. Additional data and plots generated for each regional analysis not presented below is provided in a supplemental html file.

# Methods

## Fish Taxa

Taxa range information was pulled from PISCES (Santos et al, 2014), a comprehensive database of native fish taxa ranges for California that stores range information using HUC 12 sub-watersheds as defined within the Watershed Boundary Dataset (WBD) (<https://nhd.usgs.gov/wbd.html>). Although clusters for many different subsets of native California fish taxa, including anadromous, wide ranging, narrow-ranged, and flow sensitive (as defined in Grantham et al, 2014), could be used to create regional fish assemblages, we decided that that the Flow Sensitive species subset was the most appropriate focus for determining regional fish assemblages that would relate to and be most impacted by natural or altered flow regimes. Thus we completed clustering using the Flow Sensitive species subset (see species list in Appendix B.1).

## Regional Designations

Initially in the process of development, we explored the use of standard HUC-level watershed boundaries as the method for determining regional fish assemblages. For example, a fish assemblage could be defined simply as all flow-sensitive species within a certain HUC watershed size. At small local scales (HUC12 or HUC10), this may be the simplest way to evaluate potential species of interest or indicator species for management. However, at the statewide or large regional scale, even HUC8 or HUC6 scales result in hundreds or 20+ units of analysis, respectively, with potentially large numbers of species occupying limited locations within each unit of analysis. Therefore, we decided to use a geographically based clustering approach that while fundamentally defined at a hydrologically relevant scale (HUC12), would be driven by watershed connectivity independent of the HUC scale hierarchy.

In initial trials of clustering, we found that statewide clustering did not generate useful hydrogeographic assemblages due to the small scale of the HUC12 units, which caused some clusters to cover large portions of the state and include large, disconnected taxa lists. Additionally, while clusters resembling those created using our final method were present in the statewide clusters, the southern portion of the state often went unrepresented while the northern and central portion of the state saw geographically small clusters due to higher diversity. Data output from these early iterations can be found at <https://ucd-cws.github.io/eflow-species>. We therefore sought to determine appropriate regional divisions within the state for clustering that would be large enough to encompass broad geographical areas, but not so small as to limit the usefulness of clustering. We considered using HUC6 and HUC4 units as our region boundaries, but found that the number of regions remained too high (16-20+ regions) to effectively manage for clustering. We also considered using HUC6 or HUC8 units in lieu of clustering entirely but determined that they did not align well enough with the ecology, or in the case of HUC8s, that they were too detailed to use for statewide management.

The Moyle Zoogeographic Provinces (Moyle et al, 2006) are broader in scale and have ecological and geographic relevance; however, we felt that they were too coarse, even at their most detailed subprovince level, to be used directly instead of clustering. We also considered using the broader provinces instead of creating a set of regions, but preferred to keep the coastal watersheds together and make fewer overall splits to let clustering the data determine boundaries as much as possible via clustering. We therefore defined 5 regions that were created combining complete HUC4 units so that watersheds remained connected in general geographic regions. See figure 1 for a map of the final regions used, including a comparison with the Moyle Zoogeographic Provinces. Our final regions closely resemble the Moyle Zoogeographic Provinces with the minor differences described.

The five regions used for analysis are:

* **Central Valley & West Slope Sierra**  
  All HUC4s flowing out of the Central Valley via the San Francisco Bay, including the west slope of the Sierra Nevada, the Coast Range HUCs flowing into the valley, and the Pit River system
* **North Coast**  
  All HUC4s flowing out the Klamath river, as well as all HUC4s from Klamath river past San Francisco Bay to the top of Monterey Bay along the west coast
* **South Coast**  
  All HUC4s from the Monterey Bay to the US-Mexico border along the west coast.
* **Great Basin**  
  HUC4s flowing from the eastern Sierra/Modoc Plateau into the Great Basin north of Mono Lake
* **Desert**  
  HUC4s from landlocked desert drainages within California and any great basin HUC4 that contacts California south of Mono Lake (in order to keep contiguous regions), as well as Colorado River HUC4s in the southeast of the state. \*\*Note however that due to the sparse number of fish species and fish presence in the desert, this region was not subject to clustering. Species occurring in this region should be managed for independently.

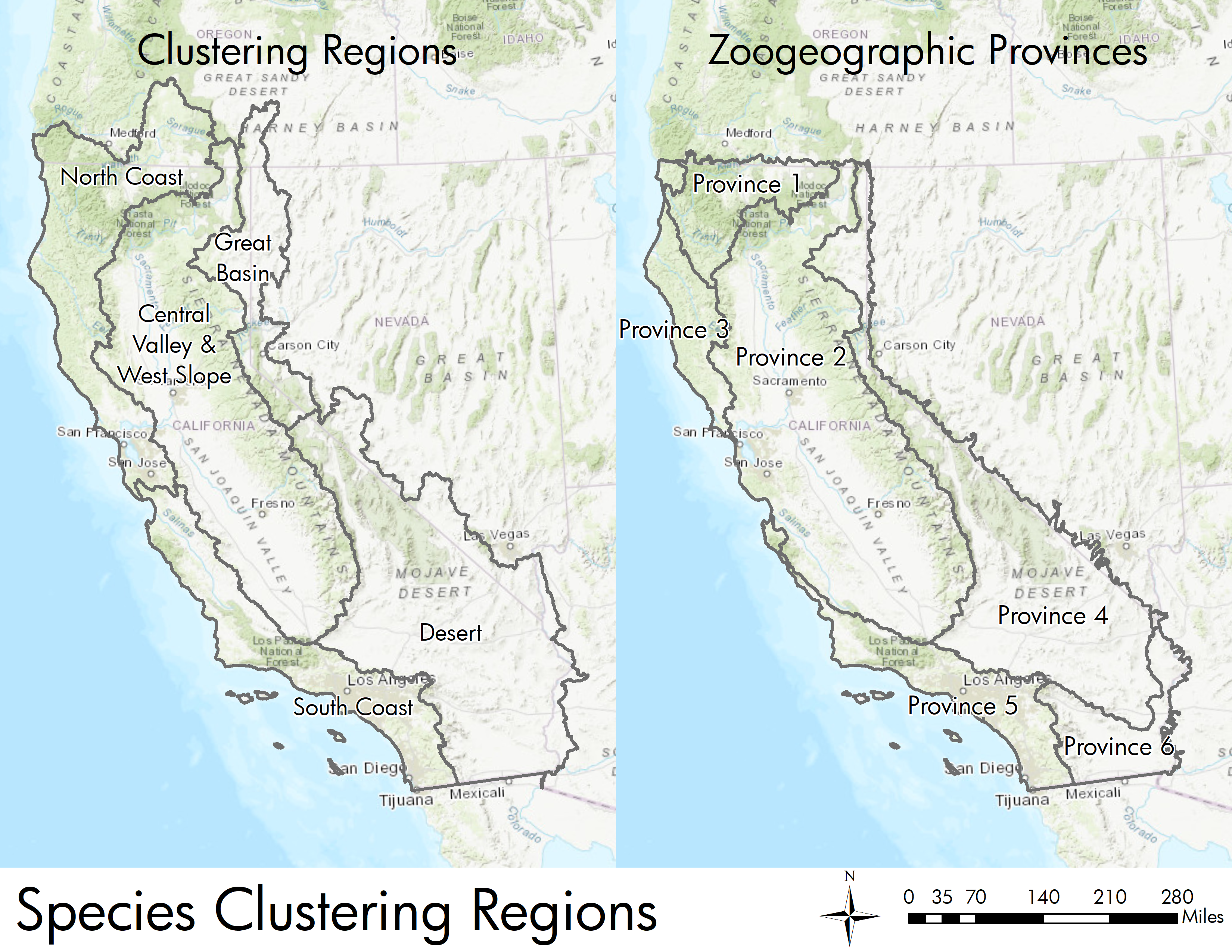


Figure : Regions created for clustering alongside the Moyle Zoogeographic Provinces

## Clustering

As part of this project, species ranges were clustered into hydrogeographic assemblages, which are sets of 3-9 spatial groupings of HUC12 level species assemblages within each region (except the desert region). Data were pulled from PISCES, using only a species’ current presence, ignoring historical presence and translocations. This data corresponded to the following data types within PISCES:

* 1: Observed current presence
* 3: Extant range - expert opinion
* 9: Reintroduction after local extirpation

Where PISCES contains records for a taxon with one of those data types, the taxon was counted as present in a given HUC12 for the purposes of the clustering. Versions using historical taxa ranges and/or translocated taxa ranges were also tried but rejected as being less useful for modern management.

Initial exploratory clustering was performed using the Kmeans tool in the GeoDa package (Anselin et al, 2006) with final clustering performed using the ArcGIS Grouping Analysis tool in the Spatial Statistics toolbox with a spatial constraint setting of “CONTIGUITY\_EDGES\_CORNERS". This constraint requires groups to be contiguous through at least one shared vertex or HUC12 edge. The clustering tool Zonation was also considered for this project, but was ultimately not chosen due to prior experience and a lack of future software support.

In development of the final methods, we created many sets of clusters with different PISCES data types, regions, and clustering settings. We ran the first round of clustering statewide with many different species groups without breaking it into regions (Figure 2). After splitting the state into regions, we focused on region refinement and determining the appropriate PISCES data types and number of groups for each region. In the next version, we used the current species presence, as well as translocated data types from PISCES. Figure 3 shows results from this clustering attempt for the Central Valley/West Slope Sierra region. This set was followed by versions that included a minimum taxa requirement for each HUC, versions without translocations, versions using historical ranges, and finally the requirement that the clusters be contiguous to make more compact clusters. With each set we evaluated the clusters for usefulness in management as discussed in the *Evaluation* section. A more complete record of these attempts, including species clusters, lists, and supporting figures is available online at <https://ucd-cws.github.io/eflow-species>

Code for analyzing the presence data and clusters and generating figures can be found at <http://github.com/ucd-cws/eflow-species>

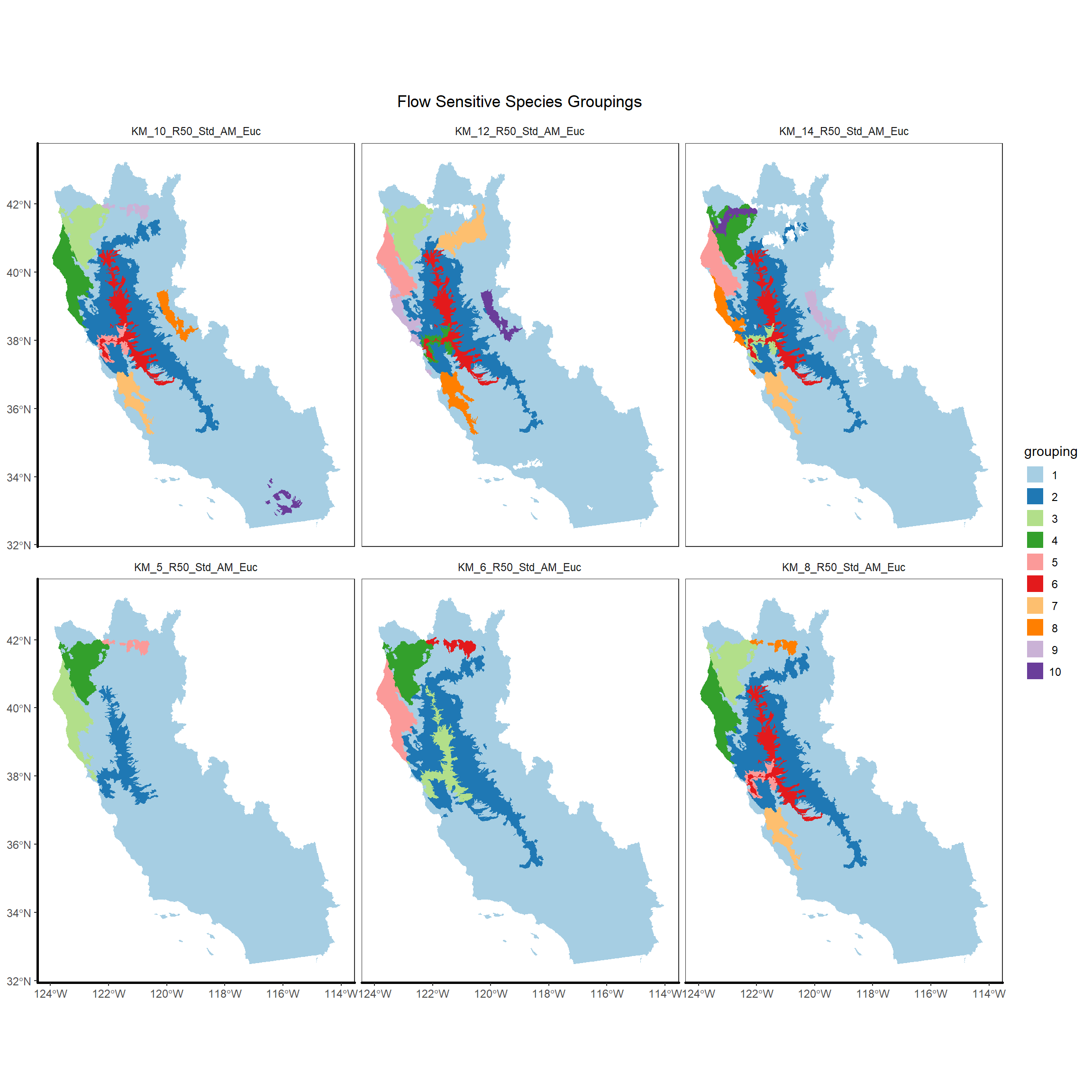


Figure 2: Initial statewide clustering of flow sensitive species without region splits. Clockwise from top left, 10 groups, 12 Groups, 14 Groups, 8 groups, 6 groups, 5 groups. Group numbers larger than 10 are shown in white.

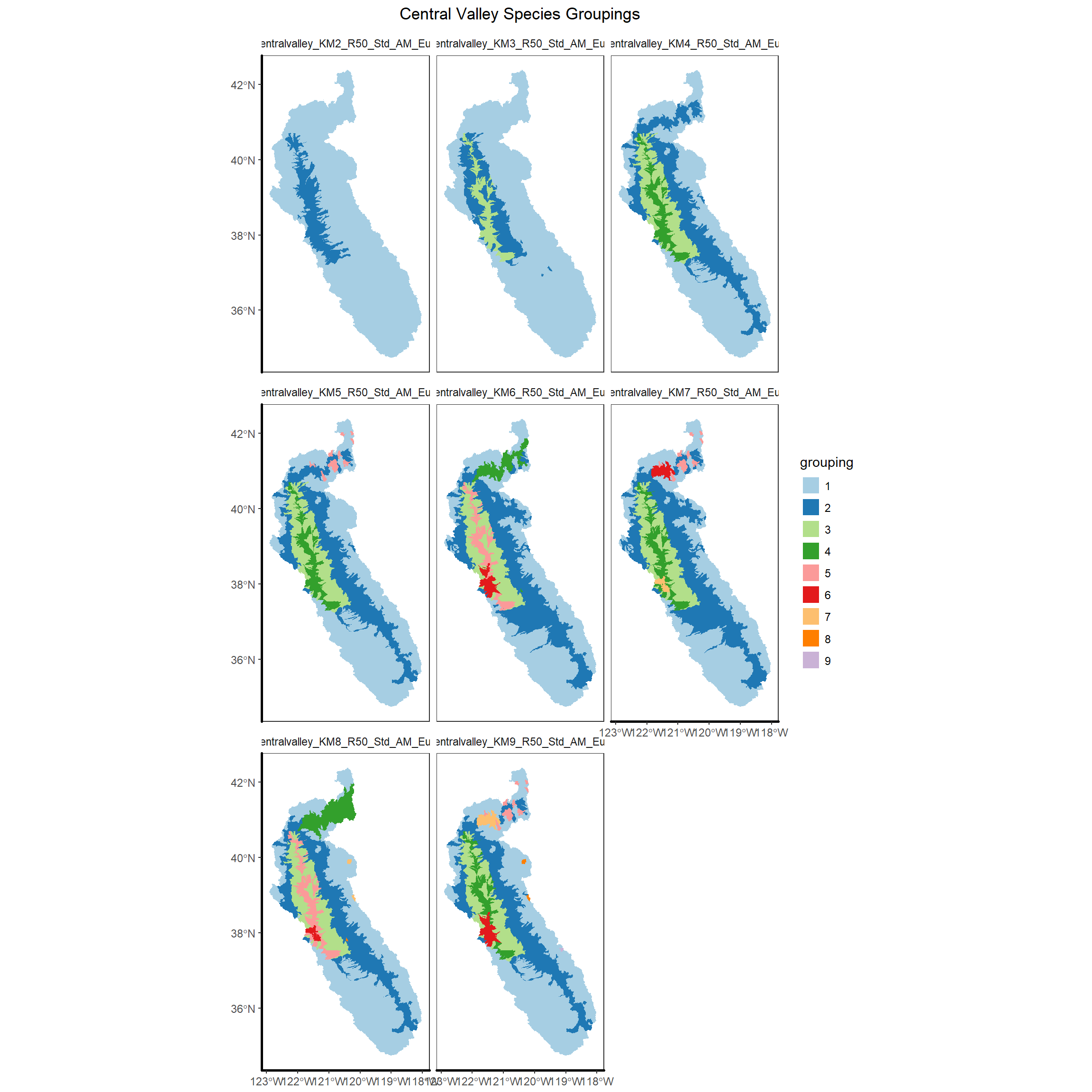


Figure 3: An early clustering for the Central Valley and West Slope, showing clusters generated when the maximum number of groups ranges from 2 to 9. Left to right from the top left, 2 Groups, 3 Groups, 4 Groups, 5 Groups, 6 Groups, 7 Groups, 8 Groups, 9 Groups.

## Evaluation

Once we determined the appropriate regional scale for clustering and the methodology for clustering as desribed above, we generated a series of clusters for each region with k set as 2 to 9. A common challenge with Kmeans-based approaches is determining the appropriate numbers of clusters to use for the data. We considered using selection criteria tied to the stream classifications as a potential way to more objectively select the appropriate number of clusters for each region. These criteria were intended to be hydrologically meaningful by trying to assess the lowest number of clusters associated with

1. the majority of stream classes present in a region were represented as a dominant class in at least one cluster
2. More than half of clusters had a dominant stream class that covered more than 60% of the cluster’s area

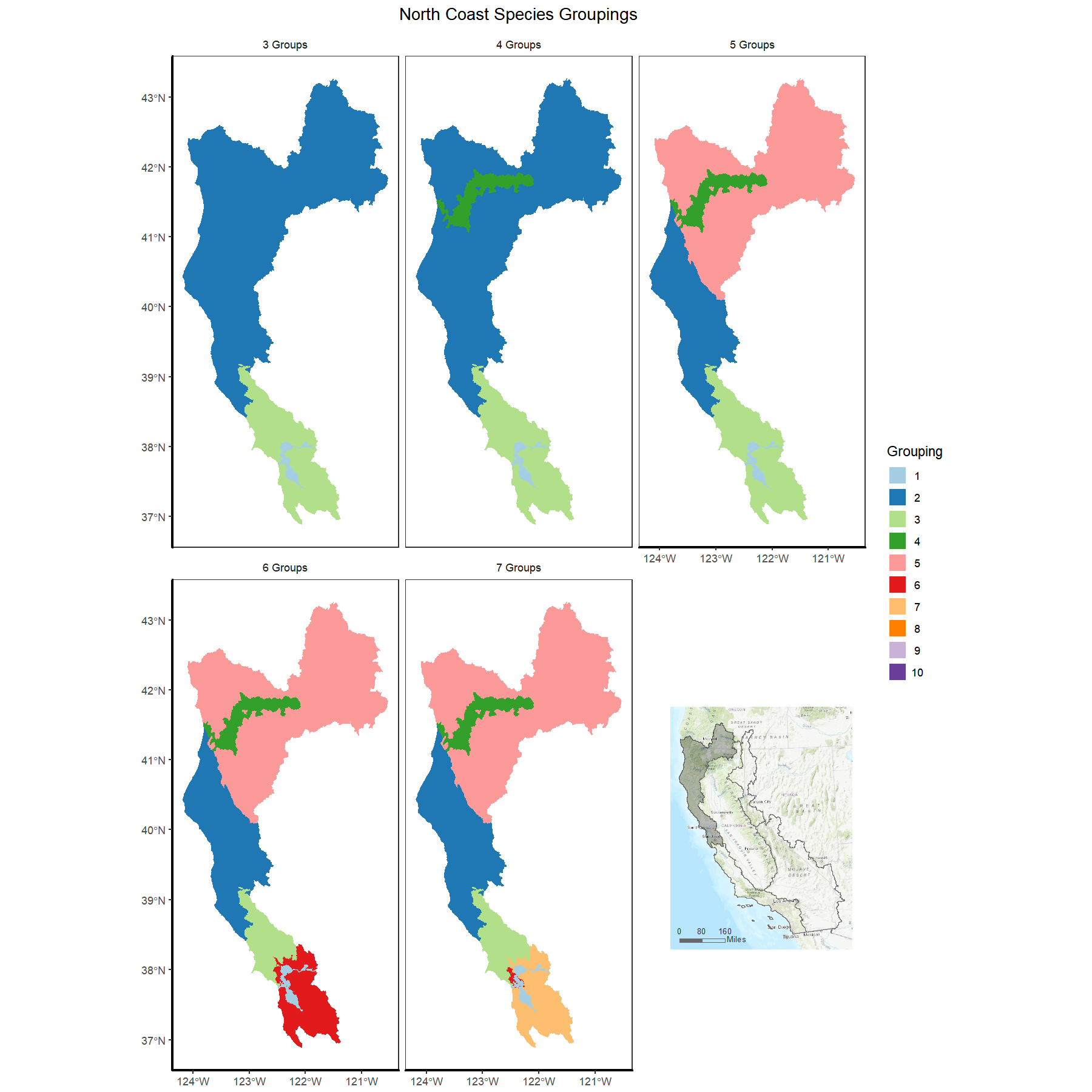
However, after repeated clustering iterations, we found that these criteria were still potentially arbitrary and not as objective as we’d hoped. Thus we chose to use expert opinion to determine what number of clusters in each region would be meaningful and useful from a management-perspective. To aid in our interpretation, we also plotted each potential group of clusters using MCA to assess which species were driving clusters and the degree of overlap between clusters. MCA is xxxxx. The final number of clusters was therefore determined via expert opinion on the ecological relevance and usefulness to management of the species composition of each cluster relative to geographic location.

# Results

## North Coast

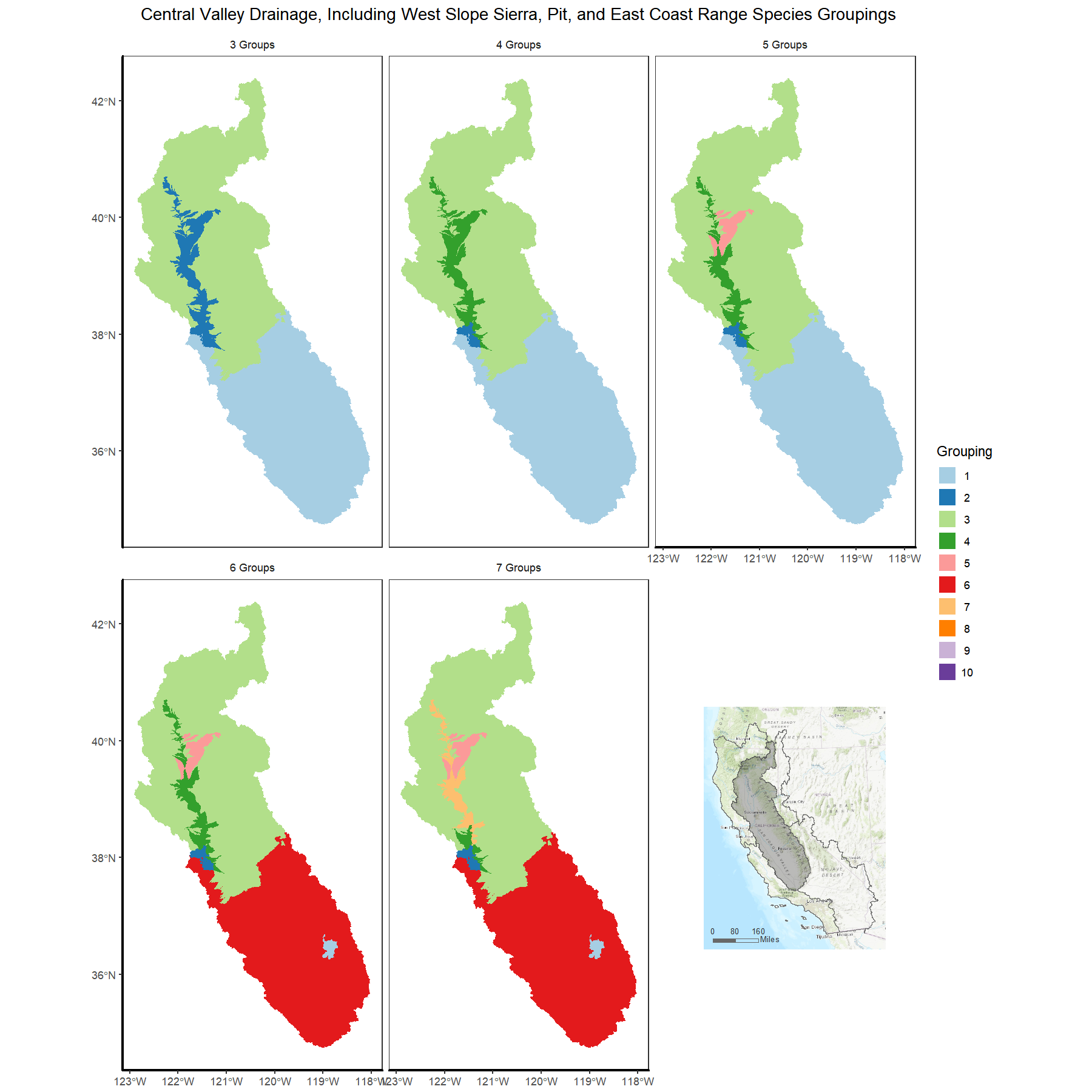
Seven potential sets of clusters were generated for the North Coast region (figure x).

<walk through a few example cluster sets and MCA plots to show thought process. Then show final selected results with key species (from powerpoint)



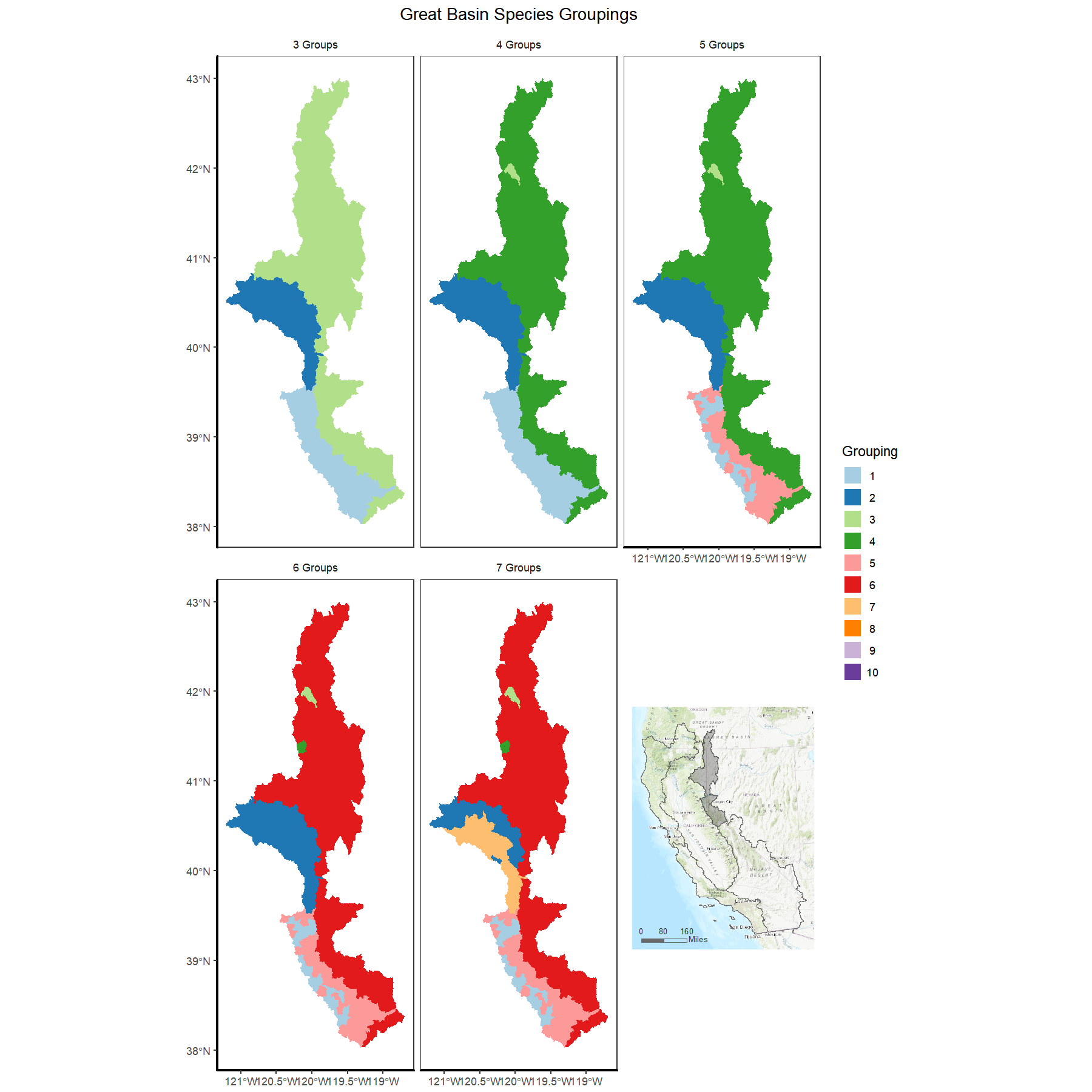
| Taxon | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
| --- | --- | --- | --- | --- | --- |
| Blue chub |  |  |  |  | + |
| California Coast fall Chinook salmon |  | + | + |  |  |
| California roach |  |  | + |  |  |
| Central California coast winter steelhead | + |  | + |  |  |
| Central Coast coho salmon |  | + | + |  |  |
| Central Valley fall Chinook salmon | + |  |  |  |  |
| Central Valley late fall Chinook salmon | + |  |  |  |  |
| Central Valley spring Chinook salmon | + |  | + |  |  |
| Central Valley winter Chinook salmon | + |  |  |  |  |
| Delta smelt | + |  | + |  |  |
| Hardhead |  | + | + |  |  |
| Inland threespine stickleback | + | + | + | + | + |
| Klamath Mountains Province summer steelhead |  |  |  | + | + |
| Klamath Mountains Province winter steelhead |  |  |  | + | + |
| Klamath River lamprey |  |  |  | + | + |
| Klamath largescale sucker |  |  |  |  |  |
| Longfin smelt | + | + | + | + |  |
| Lost River sucker |  |  |  |  | + |
| Lower Klamath marbled sculpin |  |  |  | + |  |
| Northern California coast summer steelhead |  | + |  |  |  |
| Northern California coast winter steelhead |  | + |  |  |  |
| Northern green sturgeon |  | + |  | + | + |
| Pacific lamprey | + | + | + | + | + |
| Riffle sculpin |  |  | + |  |  |
| River lamprey | + | + | + | + | + |
| Sacramento hitch | + | + | + |  |  |
| Sacramento pikeminnow |  | + | + |  |  |
| Sacramento splittail | + |  | + |  |  |
| Sacramento sucker |  |  | + |  |  |
| Sacramento tule perch | + |  | + |  |  |
| Shortnose sucker |  | + |  |  | + |
| Southern Oregon Northern California coast coho salmon |  | + |  | + | + |
| Southern coastal roach |  |  | + |  |  |
| Southern green sturgeon | + |  | + |  |  |
| Tidewater goby |  | + | + | + | + |
| Upper Klamath marbled sculpin |  |  |  |  | + |
| Upper Klamath Trinity fall Chinook salmon |  |  |  | + | + |
| Upper Klamath Trinity spring Chinook salmon |  |  |  | + | + |
| Western brook lamprey | + | + | + | + | + |
| White sturgeon | + | + | + | + | + |

## Central Valley – West Slope Sierras



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Taxon | Group 1 | Group 2 | Group 3 | Group 4 |
| Bigeye marbled sculpin |  |  | + |  |
| California roach | + |  | + | + |
| Central Valley fall Chinook salmon | + | + | + | + |
| Central Valley late fall Chinook salmon | + | + | + | + |
| Central Valley spring Chinook salmon |  | + | + | + |
| Central Valley winter Chinook salmon |  | + | + | + |
| Delta smelt | + | + | + | + |
| Hardhead | + | + | + | + |
| Inland threespine stickleback | + | + | + | + |
| Kaweah roach | + |  |  |  |
| Kern River rainbow trout | + |  |  |  |
| Kern brook lamprey | + |  |  |  |
| Longfin smelt |  | + |  |  |
| Modoc sucker |  |  | + |  |
| Northern roach |  |  | + |  |
| Pacific lamprey | + | + | + | + |
| Pit sculpin |  |  | + |  |
| Red Hills roach |  |  | + |  |
| Riffle sculpin | + |  | + | + |
| River lamprey | + | + | + | + |
| Sacramento hitch | + | + | + | + |
| Sacramento pikeminnow | + | + | + | + |
| Sacramento speckled dace |  |  | + | + |
| Sacramento splittail | + | + | + | + |
| Sacramento sucker | + | + | + | + |
| Sacramento tule perch | + | + | + | + |
| Southern green sturgeon | + | + | + | + |
| Western brook lamprey | + | + | + | + |
| White sturgeon | + | + | + | + |

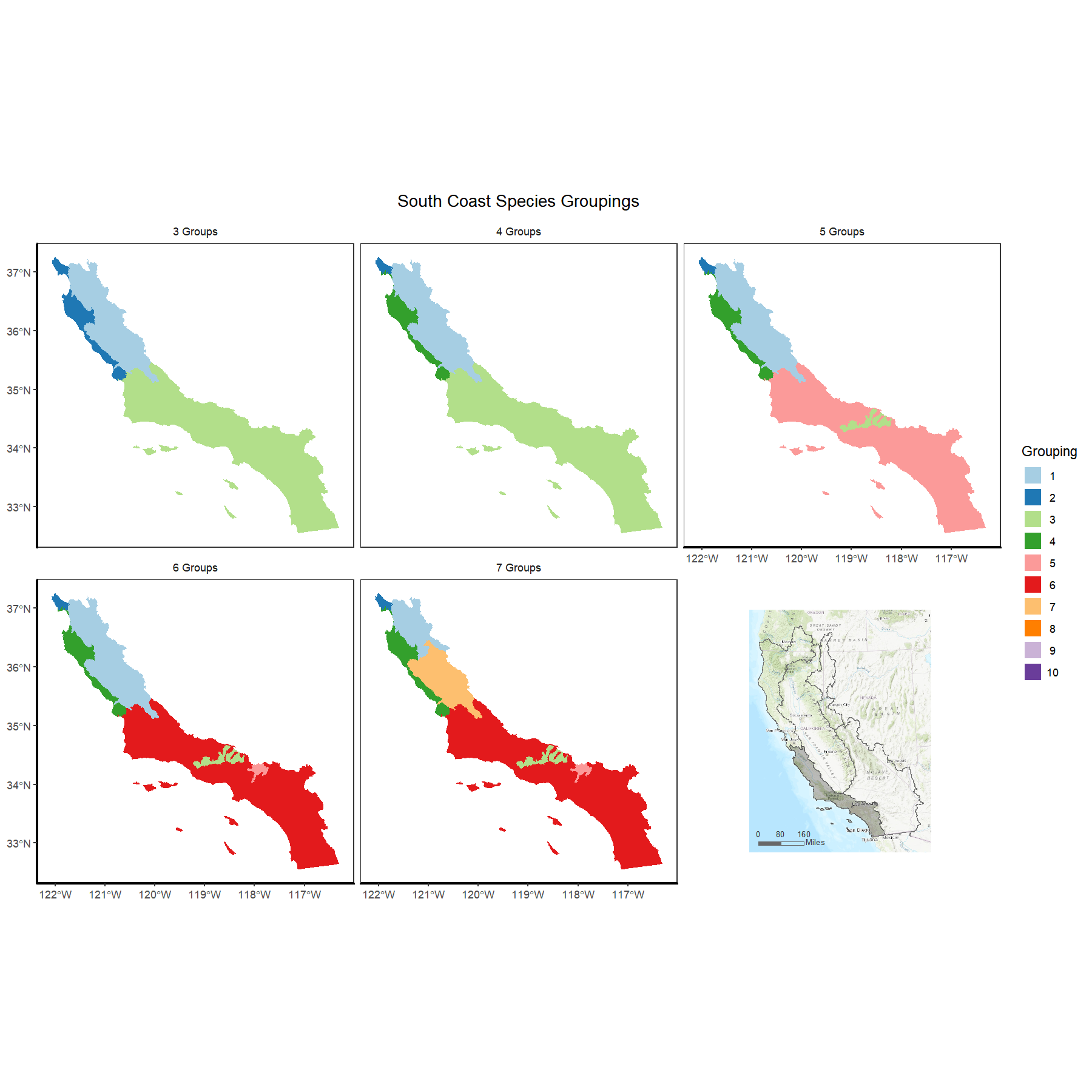
## Great Basin



|  |  |  |  |
| --- | --- | --- | --- |
| Taxon | Group 1 | Group 2 | Group 3 |
| Lahontan cutthroat trout | + |  |  |
| Lahontan mountain sucker | + | + | + |
| Lahontan speckled dace | + | + | + |
| Mountain whitefish | + |  |  |
| Sacramento speckled dace |  |  | + |

## South Coast

## 



| Taxon | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
| --- | --- | --- | --- | --- | --- |
| Arroyo chub |  |  |  |  | + |
| Central California coast winter steelhead |  | + |  |  |  |
| Inland threespine stickleback | + | + |  | + | + |
| Monterey hitch | + |  |  | + |  |
| Monterey sucker | + |  |  | + |  |
| Pacific lamprey | + | + | + | + | + |
| Riffle sculpin | + |  |  |  |  |
| Sacramento pikeminnow | + |  |  | + | + |
| Sacramento speckled dace | + | + |  | + |  |
| Santa Ana speckled dace |  |  |  |  | + |
| Santa Ana sucker |  |  |  |  | + |
| South Central California coast steelhead | + |  |  | + | + |
| Southern California steelhead |  |  | + |  | + |
| Southern coastal roach | + | + |  | + |  |
| Tidewater goby | + | + |  | + | + |
| Unarmored threespine stickleback |  |  | + |  | + |

# Appendix B.1 - Species List for Flow Sensitive Species

repaste formatted table