ArchaeoGLOBE Analysis

Nick Gauthier and Ben Marwick

Last knit on: 23 August, 2019

Analysis code for the ArchaeoGLOBE project. This notebook was used to produce all the analyses and figures in the publication. All data are downloaded from their associated online repositories prior to the analysis.

Setup

Import packages needed for analysis. We'll use packages from the `tidyverse`, such as `readr`, `dplyr`, and `ggplot2` for data import, processing, and plotting. We'll also use `mgcv` for fitting nonlinear trends to the data. We'll use the `sf` and `raster` packages to handle spatial data and plotting. The `dataverse` and `osfr` packages allows us to pull the raw survey data and precomputed analysis files from their online repositories. Finally, we'll use `patchwork` (installed from GitHub) to combine multiple ggplots in the same image.

```r
library(raster)
library(tidyverse)
library(mgcv)
library(sf)
library(ggplot2)
library(dataverse)

#install patchwork and osfr from github if needed
#devtools::install_github('thomasp85/patchwork')
library(patchwork)
# devtools::install_github('centerforopenscience/osfr')
library(osfr)
```

Data import

Download all data necessary for the analysis and import into R. By default, the code chunks that actually download the data are hidden here, so please refer to the source .rmd document for the relevant code.

Read in the latest version of the ArchaeoGLOBE database and the consensus assessment from the Dataverse repository. Refer to the source .rmd document for the code to download the shapefiles from the Dataverse repository.

```r
archaeoglobe <- read_csv('data/raw-data/ARCHAEOGLOBE_PUBLIC_DATA.tab')
consensus <- read_csv('data/raw-data/ARCHAEOGLOBE_CONSENSUS_ASSESSMENT.tab')
```

Repeat for the archaeological regions shapefile. We'll use the "simplified regions" shapefile for plotting purposes, and the original shapefile for the ArchaeoGLOBE – HYDE – kk10 comparison at the end of this notebook. Refer to the source .rmd document for the code to download the shapefiles from the Dataverse repository.

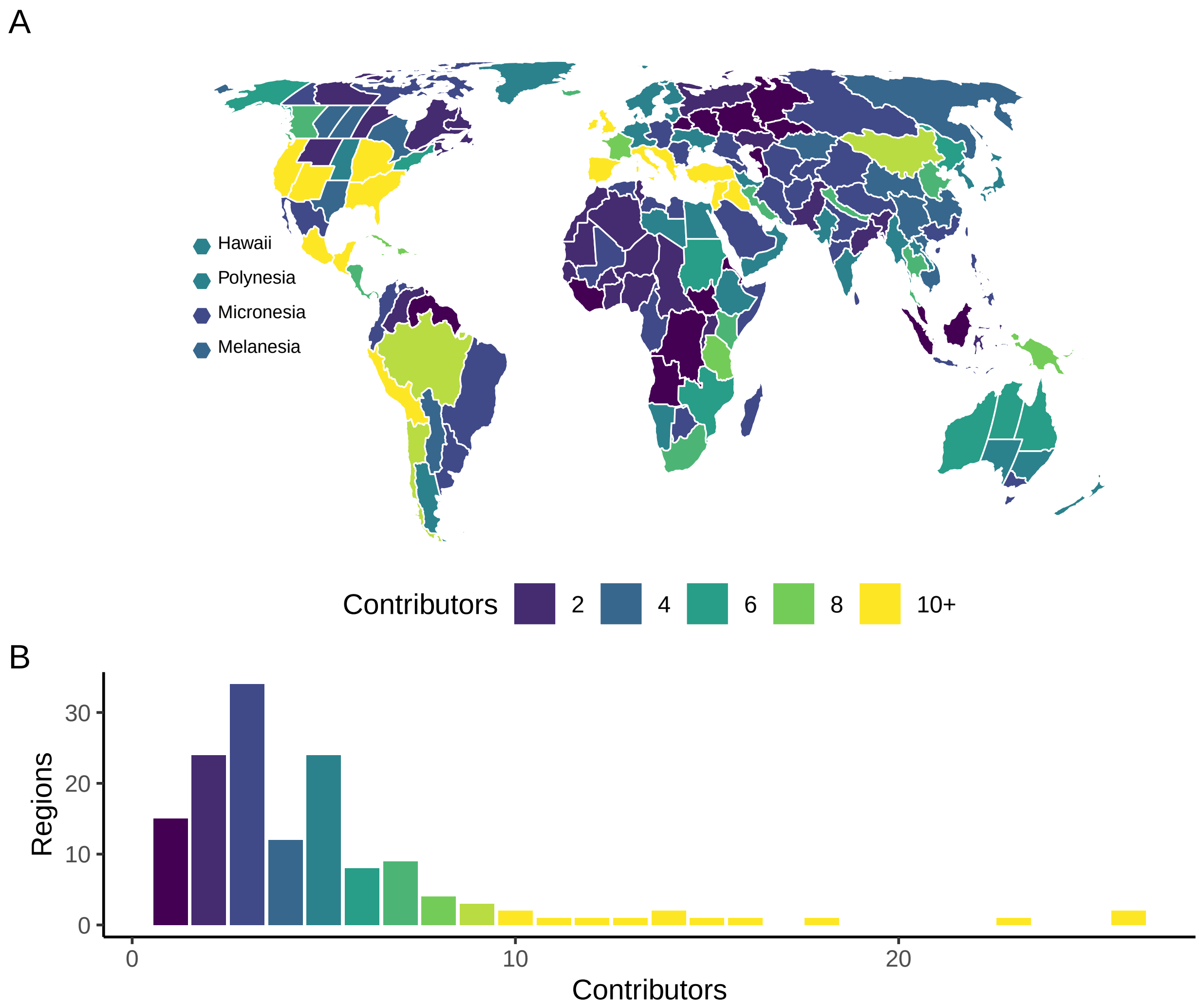
```r
# read into the current environment
regions <- st_read('data/raw-data/ArchaeoGLOBE_Simplified_Regions/ArchaeoGLOBE_Simplified_Regions.shp',
          quiet = TRUE) %>%
  # add labels for just the islands, will make plotting easier in the future
  mutate(region_label = replace(Archaeo_RG,
                                !(Archaeo_RG %in% c('Hawaii','Polynesia','Micronesia','Melanesia')), NA

regions_hyde <- st_read('data/raw-data/ArchaeoGLOBE_Regions/ArchaeGLOBE_Regions.shp',
        quiet = TRUE) %>%
  # reproject to match HYDE data
  st_transform('+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0')
```

Exploratory visualization

Before running any analyses, let's look at the data. How many responses do we have per region?

```
response_counts <- archaeoglobe %>%
  group_by(REGION_ID) %>%
  count %>%
  mutate(n10 = replace(n, n > 10, 10))
```

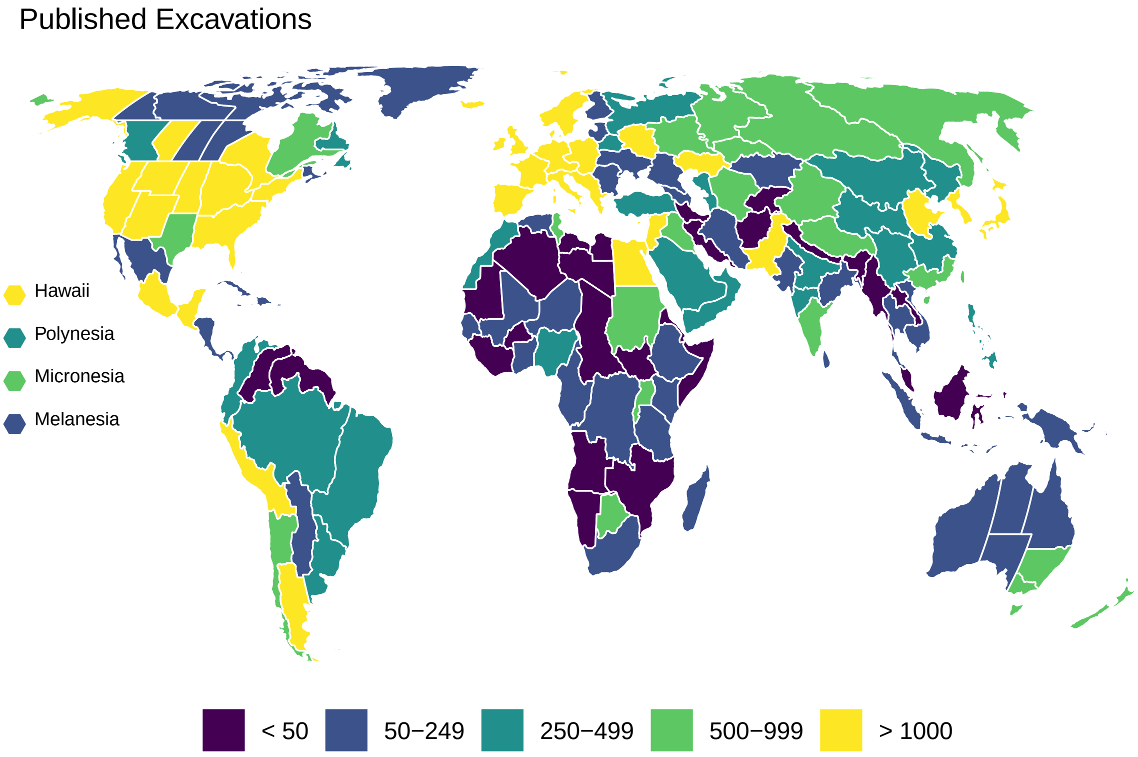


How many published archaeological excavations are estimated for each region?

```r
# a function for calculating the mode, from https://stackoverflow.com/a/46846474
calculate_mode <- function(x) {
  uniqx <- unique(x)
  uniqx[which.max(tabulate(match(x, uniqx)))]
}

# use this vector to order the RN_SITES variable by increasing number of sites
site_order <- c('< 50', '50-249', '250-499', '500-999', '> 1000')

# find the modal response for the number of published excavations in each region
site_counts <- archaeoglobe %>%
  select(REGION_ID, RN_SITES) %>%
  group_by(REGION_ID) %>%
  summarise(sites = calculate_mode(RN_SITES)) %>%
  mutate(sites = ordered(sites, levels = site_order))
```

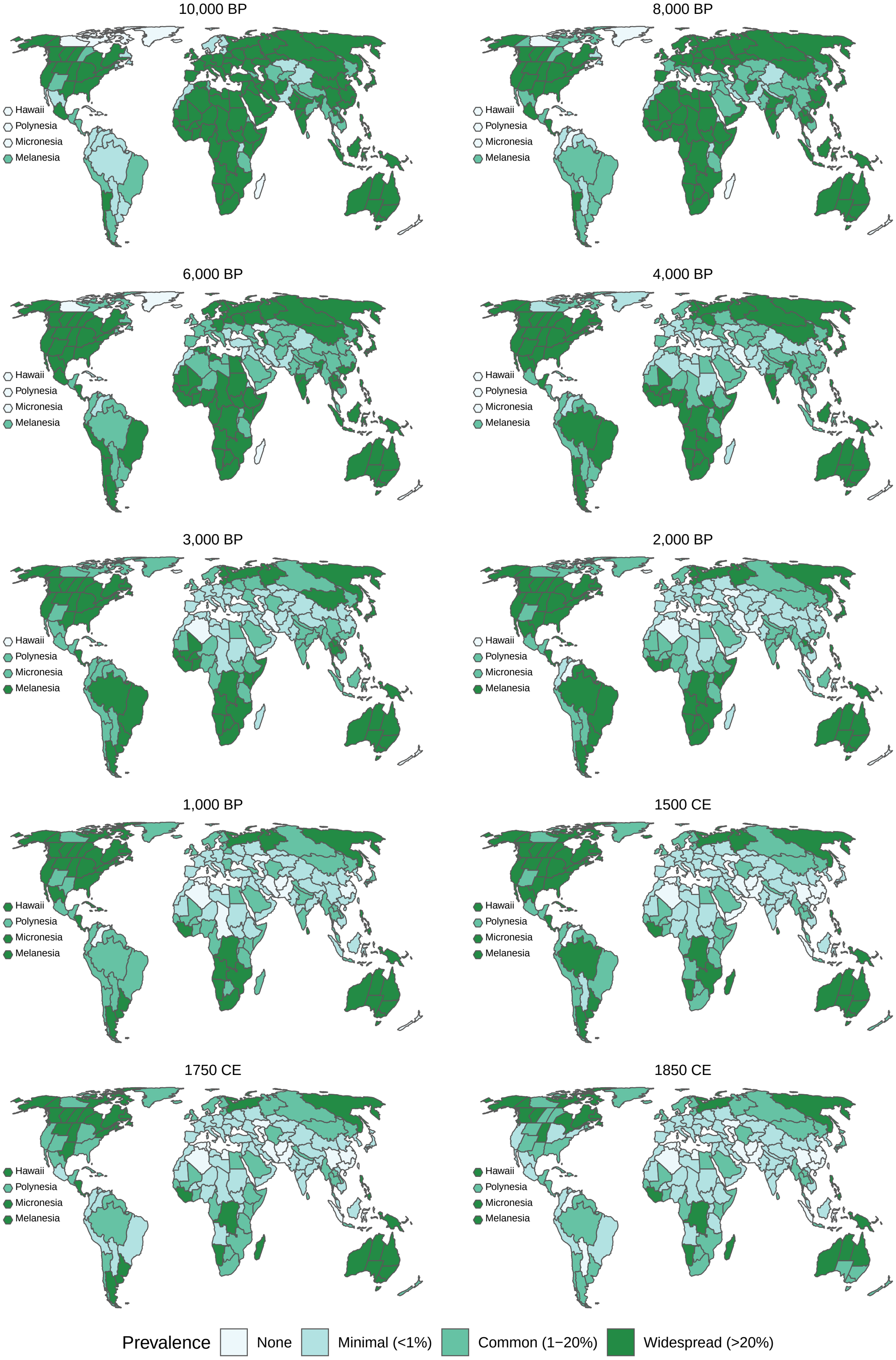


Consensus Assessment

Plot the consensus assessment data for each land-use type.

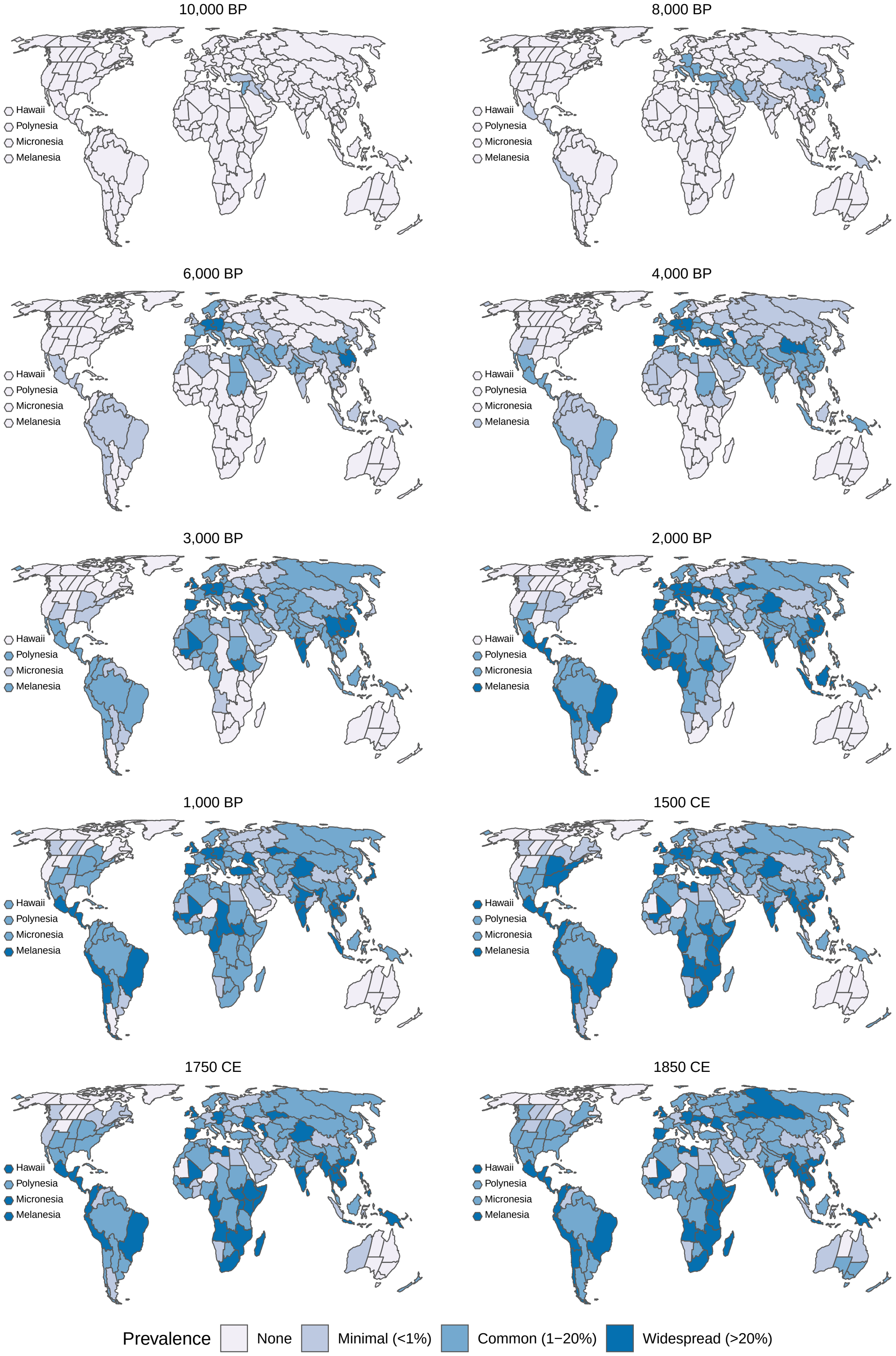
Foraging/Hunting/Gathering

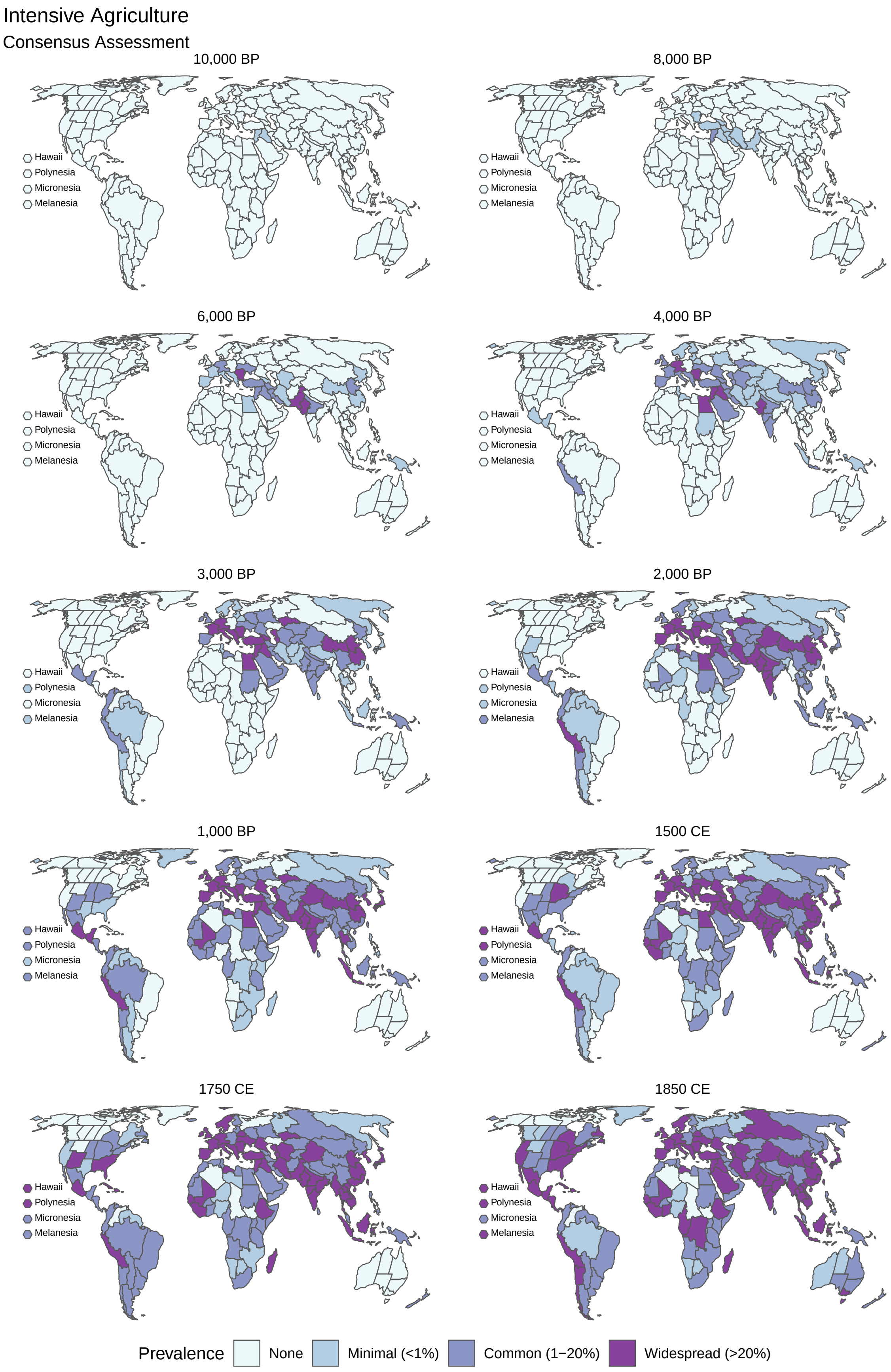
Consensus Assessment



Extensive Agriculture

Consensus Assessment





Intensive Agriculture

Consensus Assessment

10,000 BP

8,000 BP

6,000 BP

4,000 BP

3,000 BP

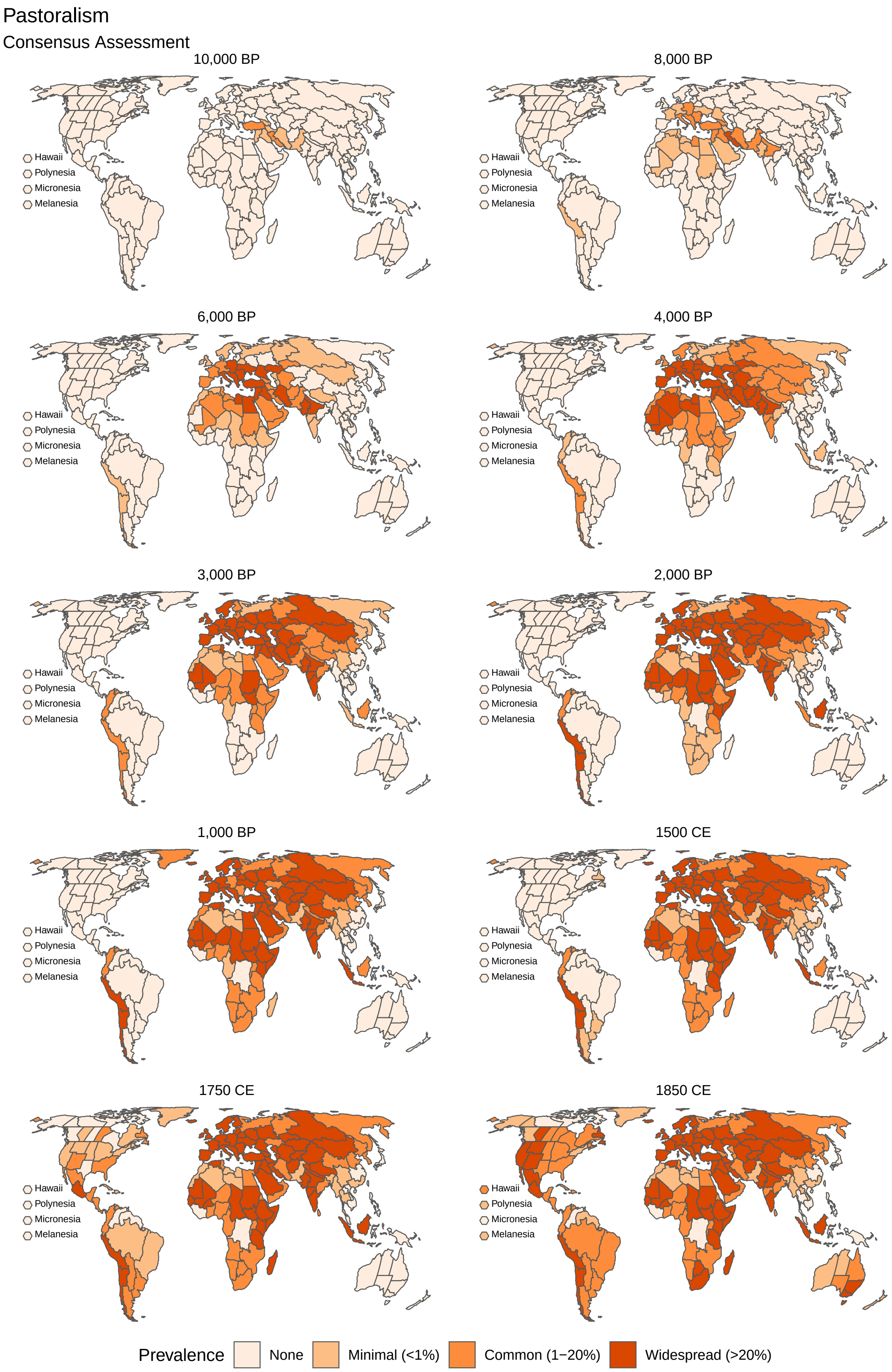
2,000 BP

1,000 BP

1500 CE

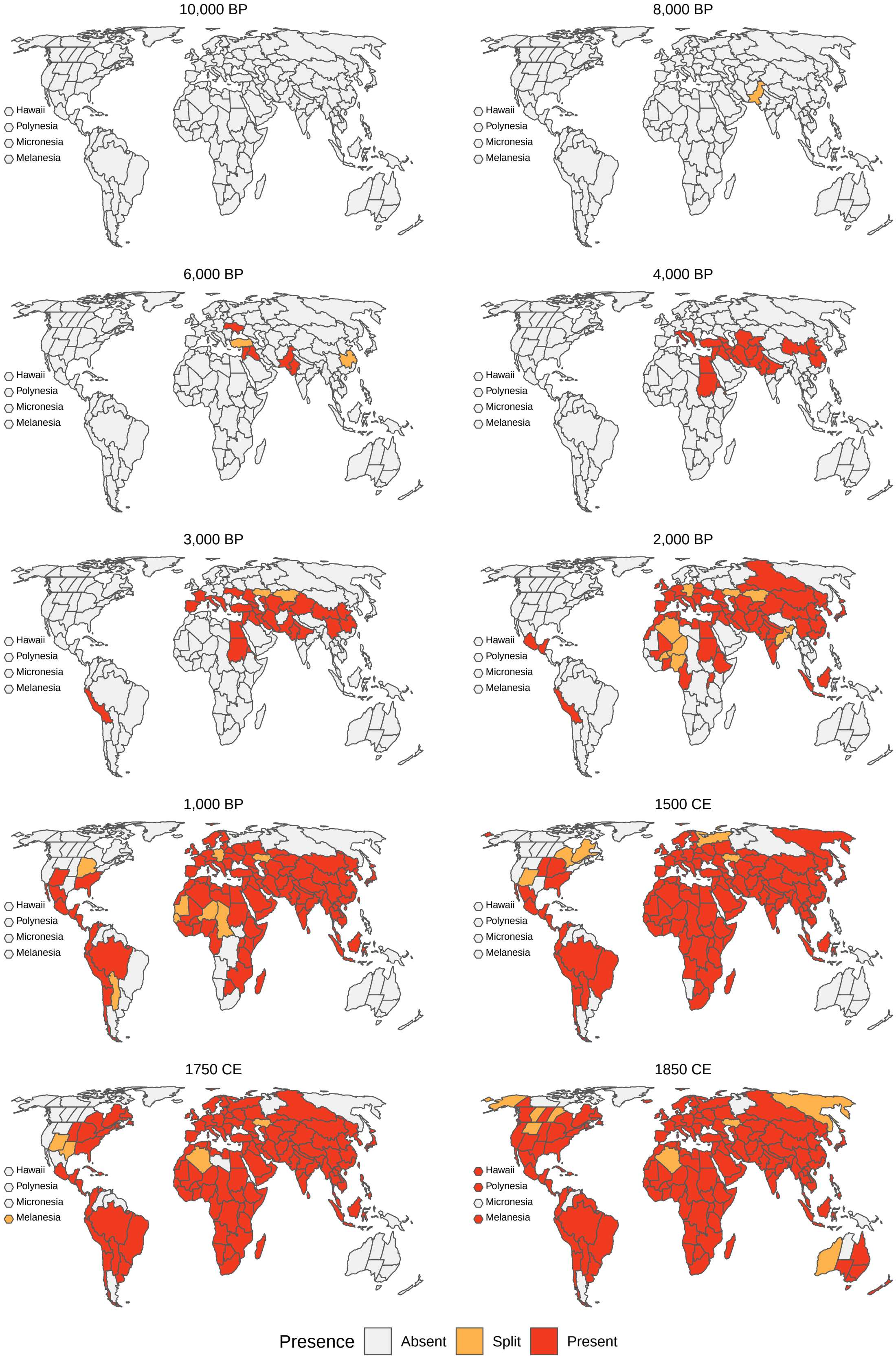
1750 CE

1850 CE


Pastoralism
Consensus Assessment
10,000 BP
8,000 BP
6,000 BP
4,000 BP
3,000 BP
2,000 BP
1,000 BP
1500 CE
1750 CE
1850 CE
Hawaii
Polynesia
Micronesia
Melanesia
Prevalence
None
Minimal (<1%)
Common (1−20%)
Widespread (>20%)

Urban Centers

Consensus Assessment



GAMM Trends

Here we use Generalized Additive Models (GAMs), a flexible form of nonlinear regression model capable of fitting smooth, time-varying trends to the ordered categorical ArchaeoGLOBE response data. We model ordered categorical data using a latent variable following a logistic distribution. The model identifies a series of cut points, which correspond the the probabilities of the latent variable falling within each of our categories.

We fit two sets of trends. One trend is fitted to all the data simultaneously, representing the global trend across all archaeological regions. Then we fit region-level trends, which represent the deviation of each region from the global trend. By penalizing the “wiggliness” of the trend lines, we allow regional trends that don’t significantly deviate from the global trend to be penalized to 0, effectively reducing that particular region to the global trend. This is a form of partial pooling, allowing the model to share information between groups and in so doing make the results less sensitive to regions with exceptionally low response rates.

After fitting the model, we can extract the region-specific trends, use a k-means clustering algorithm to group together regions with similar trends, and map the results. We repeat this analysis for both self-reported expertise and perceived data quality.

Analysis functions

Define some analysis functions that we’ll be using repeatedly in the analysis, so that we don’t have to keep copying and pasting the same lines of code.

This function subsets the data to highlight a variable of interest, and converts it from a wide to a long “tidy” format to make analysis and plotting easier.

```
preprocess <- function(prefix, categories){
  archaeoglobe %>% # start with the full ArcheoGlobe data
    # drop columns not related to the variable of interest
    select(c(CONTRIBUTR:LAND_AREA, starts_with(prefix))) %>%
    gather(time, value, starts_with(prefix)) %>% # one value per row
    mutate(time = parse_number(time) * -1, # convert time period labels to years
           value = ordered(value, levels = categories),
           cat_num = as.numeric(value)) %>%
    mutate_if(is.character, as.factor) # convert characters to factors
}
```

This function takes a data frame produced by the above function and fits GAM to the global trend and local deviations for each region, accounting for inter-observer variability. This function takes as arguments a preprocessed data frame containing time slices, regions, contributors, and the ordered categorical response variable transformed to a numeric vector.

```
cores <- max(parallel::detectCores() / 2, 1) # physical cores for parallelization
cl <- parallel::makeCluster(cores)

fit_gam <- function(x, n_cats){
  bam(cat_num ~
        # this spline is for the global trend
        s(time, bs = 'cr', m = 2) +
        # region-specific trends. bs = 'ts' and m = 1
        # help penalize deviation from the global model
        s(time, by = REGION_LAB, bs = 'cs', m = 1) +
        # add back in region-specific intercepts
        REGION_LAB  +
        # model contributor as a random effect
        s(CONTRIBUTR, bs = 're'),
      data = x, # data frame to analyze
      family = ocat(R = n_cats), # ordered categorical with n levels
      # final 3 arguments just speed up the model fitting
      method = 'fREML',
      discrete = TRUE,
      cluster = cl)
}
```

This function extracts the estimated trends for each region, incorporating the global and regional splines as well as the region and contributor specific intercepts. Then it clusters these trends into 6 discrete clusters using k-means. The choice of 6 clusters is somewhat arbitrary, and is made simply based on visual comparisons of different cluster solutions with the goal of ensuring visually interpretable results.

```
extract_trends <-function(mod, n_clusters = 6){
  set.seed(1000) # set seed for reproducability of clusters
  archaeoglobe %>% # create dummy data for prediction in the following lines
    select(REGION_LAB) %>%
    group_by(REGION_LAB) %>%
    slice(1) %>%
    slice(rep(1:n(), each = 198)) %>%
    ungroup %>%
    mutate(time = rep_len(seq(-10000, -150, 50), n()),
           CONTRIBUTR = 'CYRBU') %>% # select an arbitrary contributor
    mutate(preds = predict(mod, .)) %>% # estimate trend lines
    mutate(preds = plogis(preds)) %>% # transform responses to [0,1] scale
    spread(time, preds) %>%
    # next is the actual kmeans clustering code
    mutate(cluster = kmeans(.[,-c(1,2)], n_clusters, iter.max = 100, nstart = 100)$cluster)
}
```

Analysis

Now we use the functions defined above to estimate trends in ArchaeoGLOBE data. For convenience, first define a data frame that lists the prefixes of the variables we are interested in (e.g. “EXP” for expertise) and the levels of the ordered factors associated with each variable. This will make it easier to quickly focus on a specific variable. The `tribble` command is simply a way to make a data frame by row rather than column, which makes the code easier to read.

```
response_levels <- tribble(
  ~prefix, ~categories,
  'EXP', c('None', 'Low', 'High'),                  # Expertise
  'DQ', c('Unknown', 'Low', 'Moderate', 'Good'),    # Data Quality
  'HUNT', c('none', 'minimal (<1%)', 'common (1-20%)', 'widespread (>20%)'),
  'EXAG', c('none', 'minimal (<1%)', 'common (1-20%)', 'widespread (>20%)'),
  'INAG', c('none', 'minimal (<1%)', 'common (1-20%)', 'widespread (>20%)'),
  'PAST', c('none', 'minimal (<1%)', 'common (1-20%)', 'widespread (>20%)'),
  'URBN', c('Absent', 'Present')
)
```

Now map each of the above functions to each variable. This allows us to run the analysis for all variables of interest in a single step, and save all the outputs in a tibble format for easy plotting. If you’re running this for the first time, it should take about 40 minutes to run on a Intel NUC with a 5th-gen Intel Core i7-5557U processor and 16gb of RAM running Linux. By default, we pull pre-computed results from a repository rather than running the time consuming analysis.

```
trend_dat <- response_levels %>%
  mutate(data = map2(prefix, categories, ~preprocess(.x,.y)),
         n_cats = map_dbl(categories, length),
         mod = map2(data, n_cats, fit_gam),
         trends = map(mod, extract_trends))
```

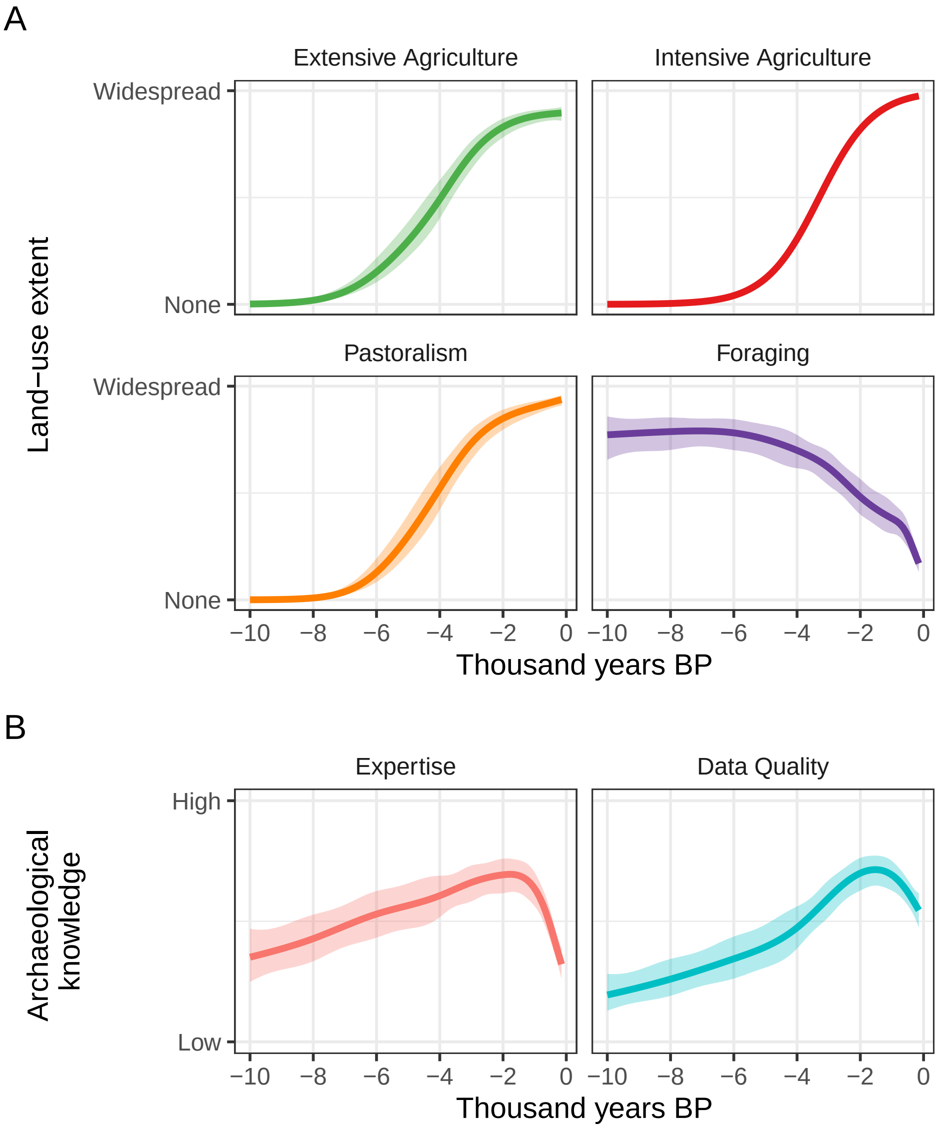
Results

First we plot out the global trends for each land use type, and compare them to the consensus estimates. Then we plot the local (regional trends) for all land use types, and map out their associated clusters. Please refer to the .rmd source file for the code to make the plots.

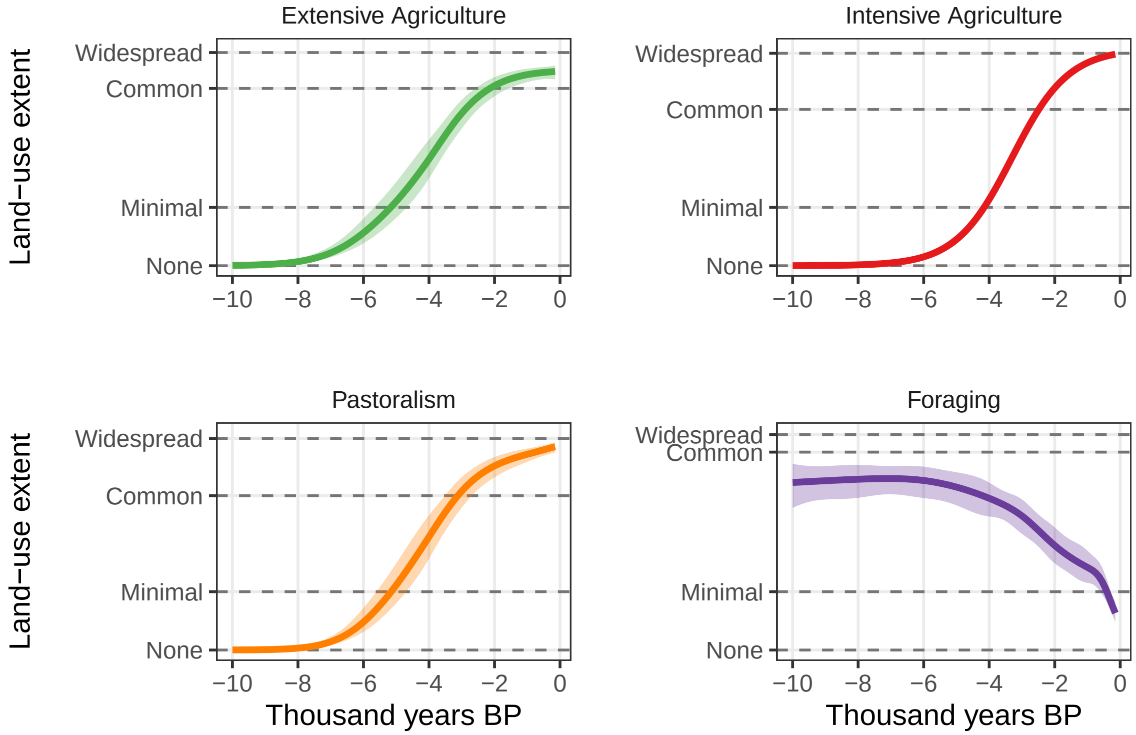
Global Trends

The global trend in foraging shows constant high prevalence until around 6,000 years ago, after which there is a smooth decline until the present day when it is very rare. Mapping out the clusters reveals a clear east-west divide, which regions in Afro-eurasia seeing foraging earlier then the global mean, and regions in the Americas and Oceania seeing later peaks in foraging.

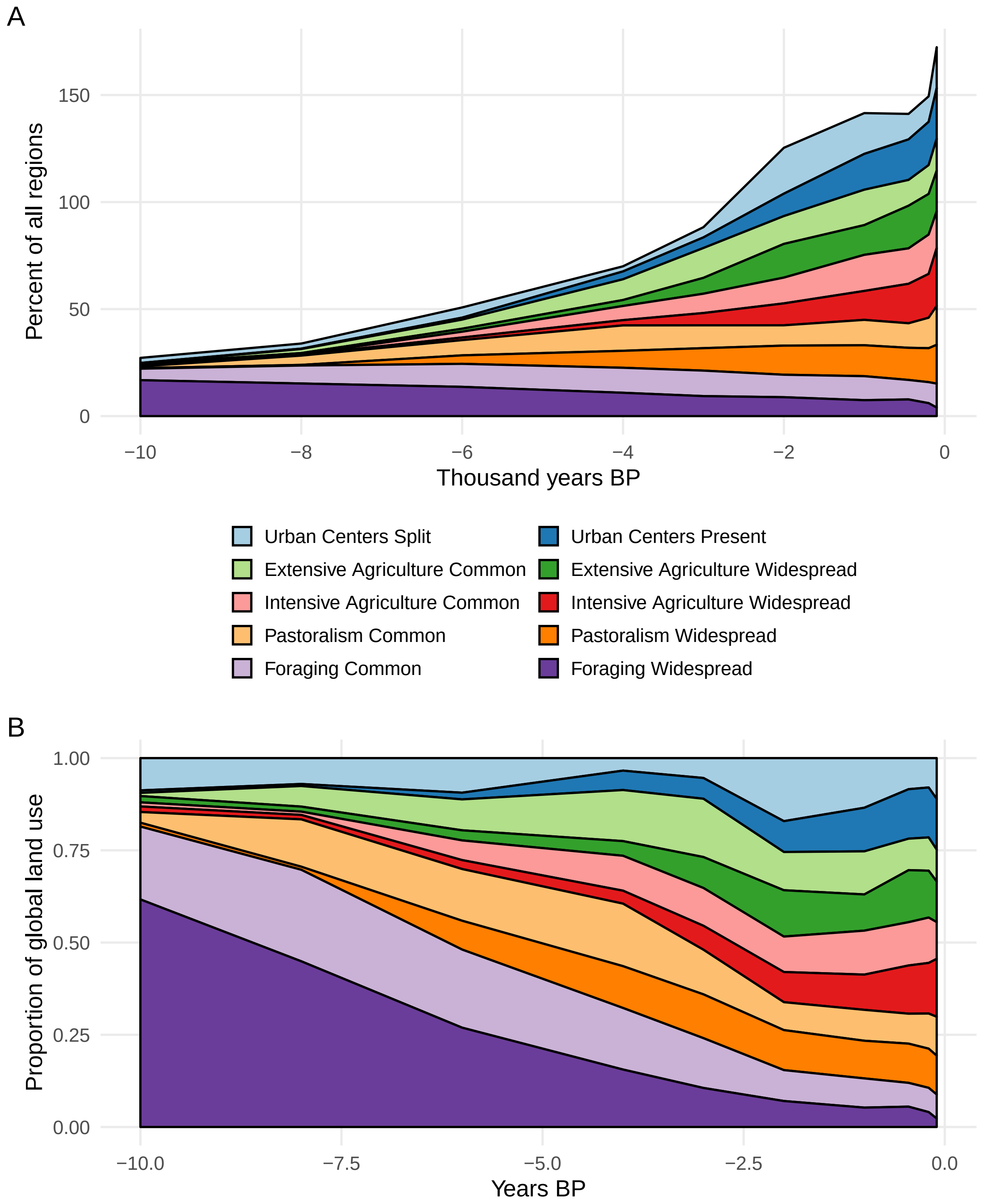
The global trends in the prevalence of pastoralism, extensive and intensive agriculture, and urbanism all follow a sigmoidal curve, which means the trend is linear on the scale of the linear predictor (the ordered categorical GAM uses a logit transform as a latent link function). This means that there is a simple increase in the probability of each land use type being prevalent over time.



The numerical cutpoints between the ordered categorical response levels estimated by the model vary across land-use type. This is a normal result of the ordered categorical regression, and basically means that different sources of error/uncertainty impact how contributors translate their mental models of areal extent (the latent, “real” value the regression is trying to estimate) into discrete categories across the different land use types.

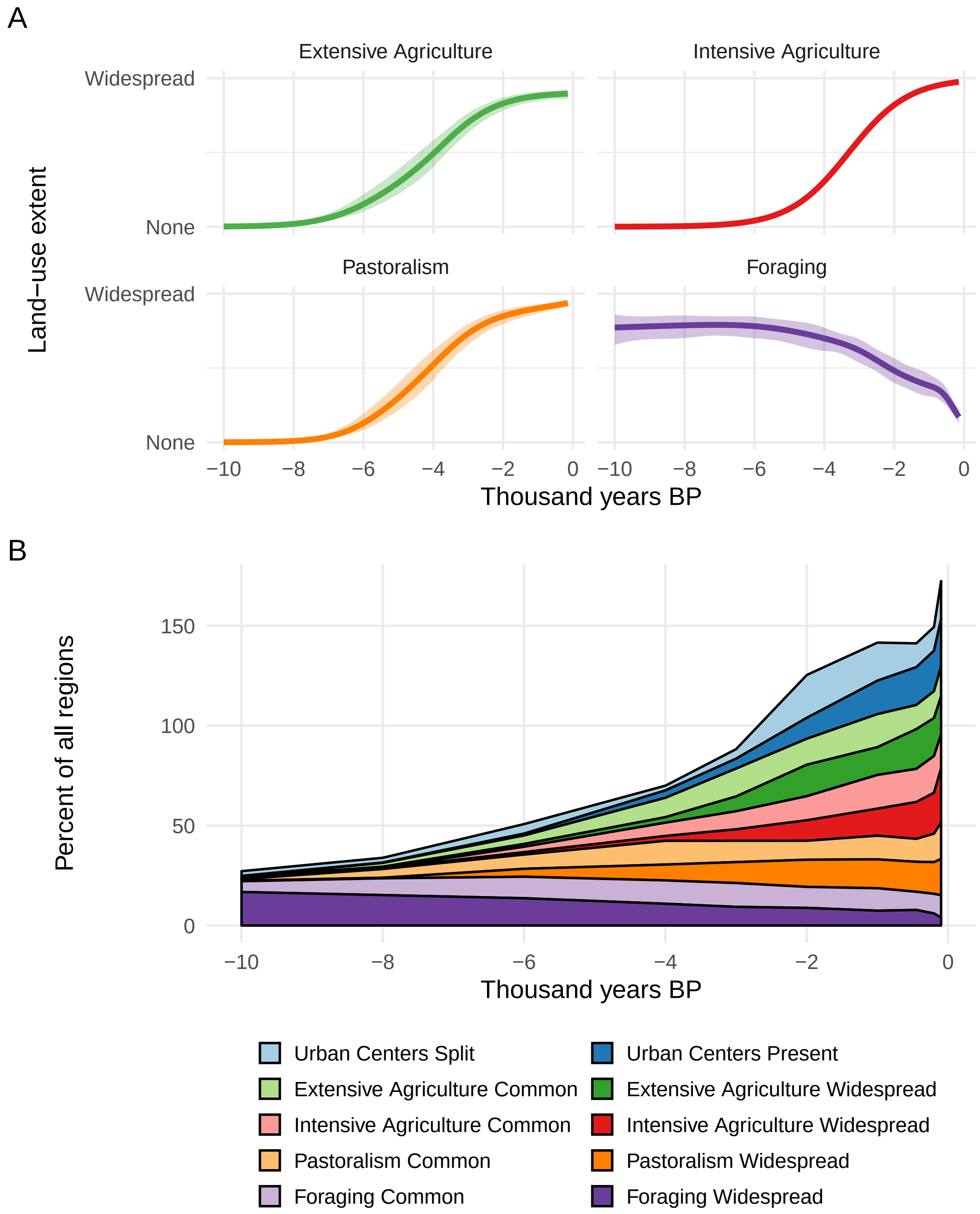


Compare the global trends to the consensus assessments. To do this, we first need to calculate the cumulative changes in land-use extent for each type.



Now plot the two analyses (GAM and consensus) together.

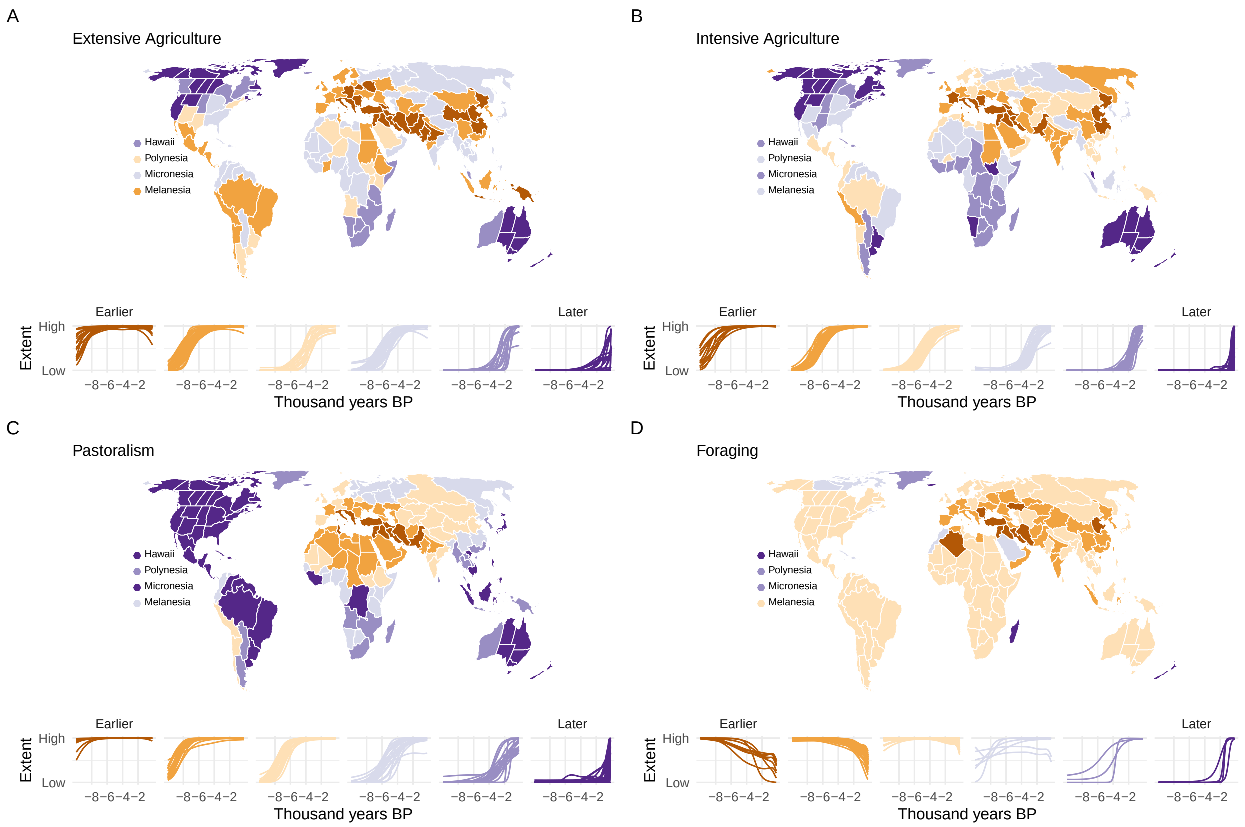
```
(g1 + theme_minimal()) / cs1 + plot_annotation(tag_levels = 'A')
```



```r
ggsave('figures/3_trends_global.png', height = 7.5, width = 7)
```

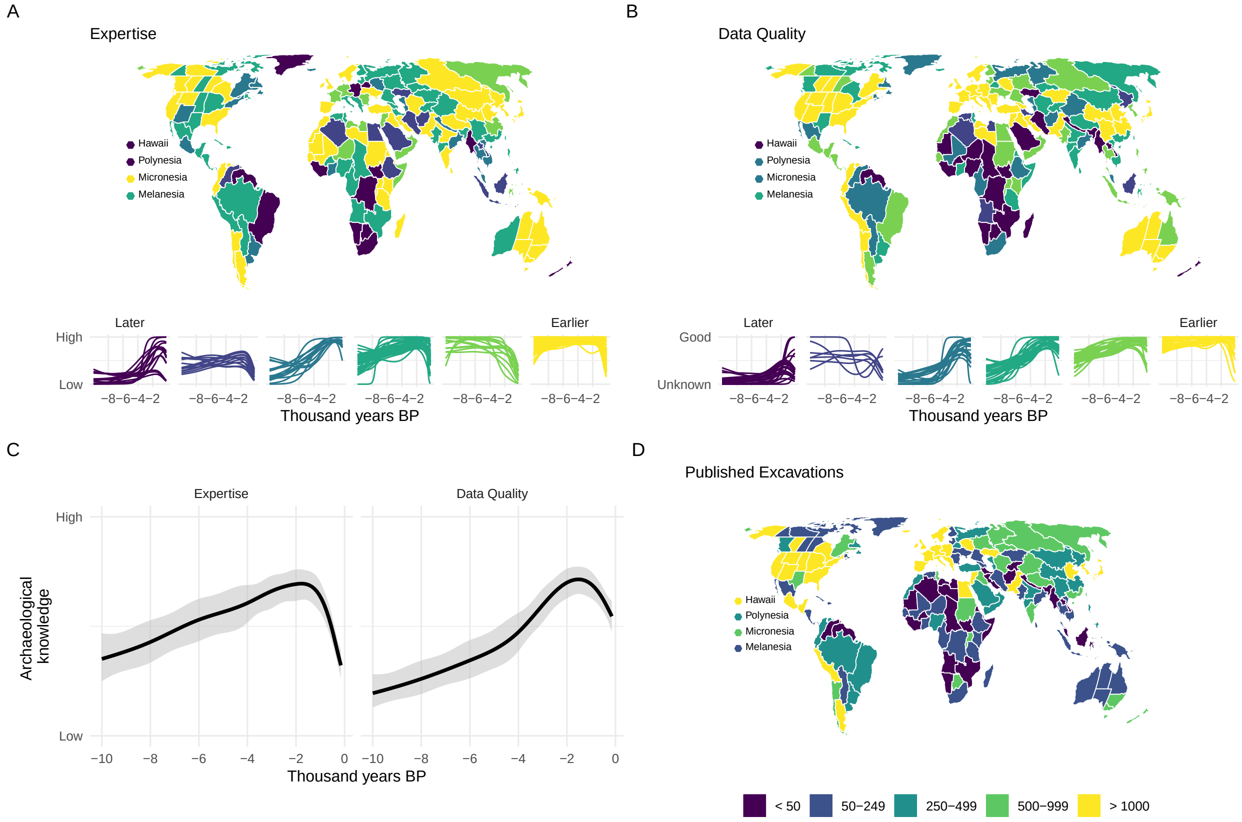
Regional Trends

Now we can break down the land-use trends by region.



How does self-professed level of expertise vary in each region over time? The global trend is a roughly linear increase in self-reported expertise from 10ka BP up to 2ka BP, then a falloff continuing to the present day. The present day expertise values are approximately the same as at 10ka BP. This makes sense, as it points to both the increased frequency of preserved archaeological materials with time as well as the reduction in archaeological attention in periods with extensive historical records.

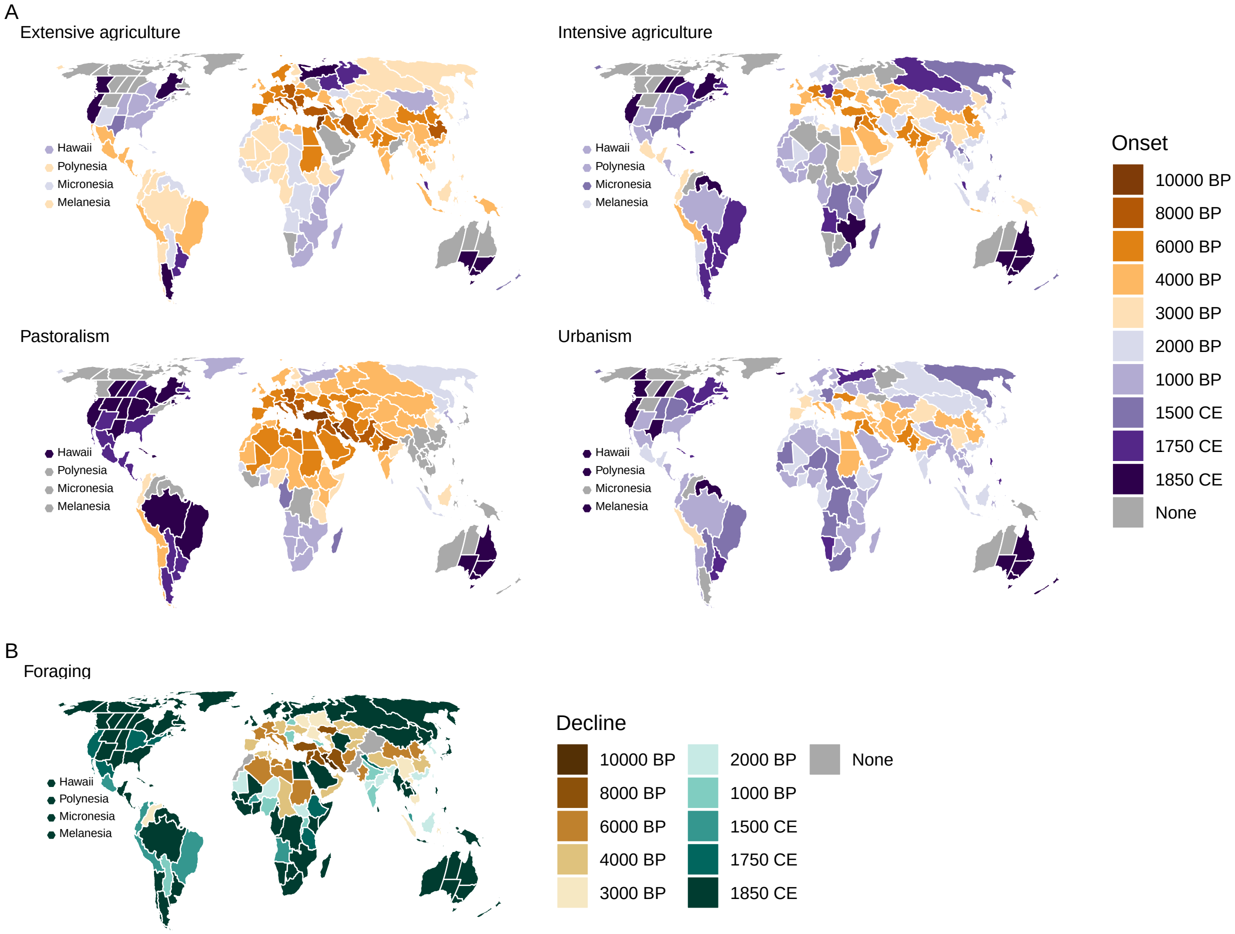
The global trend in data quality is more or less the same as the expertise data, with the peak in data quality occurring more recently than for expertise and with a less dramatic falloff leading to the present day. Unlike expertise, which reaches the same values at 10ky BP and present, data quality in the present day remains high in spite of the falloff in the last 2 millennia. Also note the confidence interval for the global trend is generally wider than for the expertise responses.



Onset and decline timing

Next we visualize the timing of the onset of extensive and intensive agriculture, pastoralism, and urbanism for each region, as well as the timing of the initial decline in foraging.

```
consensus_transitions <- consensus %>%
  # get consensus data into a form to calculate onset timings
  select(Region, Label, FHG_10KBP:URBAN_1850CE) %>%
  gather(time_step, level, FHG_10KBP:URBAN_1850CE) %>%
  separate(time_step, c('type', 'time_step')) %>%
  mutate(time = parse_number(time_step),
         time = case_when(time == 1500 ~ .45,
                          time == 1750 ~ .2,
                          time == 1850 ~ .1,
                          time <= 10 ~ time),
         time = time * -1,
         type = fct_relevel(type, 'EXAG', 'INAG', 'PAS', 'URBAN', 'FHG'),
         type = fct_recode(type, `Extensive agriculture` = 'EXAG',
                          #some EXAG entries are spelled EXAGR, this fixes that
                          `Extensive agriculture` = 'EXAGR',
                          `Intensive agriculture` = 'INAG',
                          `Pastoralism` = 'PAS',
                          `Urbanism` = 'URBAN',
                          `Foraging` = 'FHG')) %>%
  filter(level %in% c('Common', 'Widespread', 'Present')) %>%
  group_by(Region, type) %>%
  summarise(onset = min(time), # calculate onset timing
            decline = max(time)) %>% # calculate decline timing
  ungroup %>%
  # join to region data for plotting
  left_join(regions, ., by = c('Archaeo_ID' = 'Region')) %>%
  # join to regions one more time for NA regions
  complete(Archaeo_ID, type) %>%
  select(Archaeo_ID, type, onset, decline) %>%
  left_join(regions, .)
```



HYDE Comparison

Here we compare the onset times for intensive agriculture derived from the ArchaeoGLOBE consensus estimates with those calculated from HYDE 3.2 and KK10 data. We assess these differences at both the Common (>= 1%) and Widespread (>= 20%) levels.

Download and import HYDE 3.2 and KK10 land-use reconstructions. Refer to the source .rmd document for the code to download the respective .zip and .nc files from their ftp servers. By default, we pull pre-computed results from a repository rather than running the time consuming download. If you would like a fresh download of the data, refer to the .rmd source file for the necessary code, which will download about 500mb and 18gb to your computer for each dataset, respectively.

```
hyde <- list.files('data/raw-data/HYDE', full.names = TRUE) %>%
  .[c(11:8, 3, 1:2, 4:7)] %>% #  temporal order
  map(raster) %>%
  brick %>%
  `crs<-`(value = '+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0')

kk10 <- brick('data/raw-data/KK10.nc') %>%
  # select time slices of interest
  .[[c(51, 2051, 4051, 5051, 6051, 7051, 7551, 7801, 7901)]] %>%
  `crs<-`(value = '+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0')
```

First calculate the total cropland area from HYDE and KK10 in each of the ArchaeoGLOBE regions, and convert to proportion of land area under cultivation in order to compare HYDE and ArchaeoGLOBE. Start by extracting the HYDE and KK10 data within each of the ArchaeoGLOBE regions. Be warned the raster::extract() command may take a while to run.

```
hyde_crop_prop <- hyde %>%
  raster::extract(regions_hyde, na.rm = TRUE, fun = sum, df = TRUE) %>% # sample at region locations
  `names<-`(c('ID', -10, -8, -6, -4, -3, -2, -1, -0.45, -0.2, -0.1, 0.05)) %>%
  gather(time, value, 2:12) %>%
  mutate(time = as.numeric(time)) %>%
  left_join(regions, by = c('ID' = 'Archaeo_ID')) %>%
  mutate(prop = value / Land_Area)

kk_anthro_prop <- kk10 %>%
  area %>%
  `*`(kk10) %>%
  raster::extract(regions_hyde, na.rm = TRUE, fun = sum, df = TRUE) %>% # sample at region locations
  `names<-`(c('ID', -8, -6, -4, -3, -2, -1, -0.45, -0.2, -0.1)) %>%
  gather(time, value, 2:10) %>%
  mutate(time = as.numeric(time)) %>%
  left_join(regions, by = c('ID' = 'Archaeo_ID')) %>%
  mutate(prop = value / Land_Area)
```

Next calculate the earliest onset time for intensive agriculture at common and widespread thresholds for the HYDE and KK10 data. Note that HYDE and KK10 are estimating different quantities (cropland vs anthropogenic land use), so the results won't be entirely consistent.

```
# data wrangling to extract onset times at different thresholds
hyde_onset <- hyde_crop_prop %>%
  filter(prop >= 0.01) %>%
  mutate(level = if_else(prop >= 0.2, 'Widespread', 'Common')) %>%
  group_by(ID, level) %>%
  summarise(onset = min(time)) %>%
  spread(level, onset) %>%
  mutate(Common = if_else(is.na(Common), Widespread, Common)) %>%
  gather(level, onset, Common:Widespread) %>%
  mutate(source = 'HYDE',
         level = if_else(level == 'Common',
                         'Common (> 1% land area)',
                         'Widespread (> 20% land area)')) %>%
  rename(Region = ID)
```

```
# same for KK10 data
kk10_onset <- kk_anthro_prop %>%
  filter(prop >= 0.01) %>%
  mutate(level = if_else(prop >= 0.2, 'Widespread', 'Common')) %>%
  group_by(ID, level) %>%
  summarise(onset = min(time)) %>%
  spread(level, onset) %>%
  mutate(Common = if_else(is.na(Common), Widespread, Common)) %>%
  gather(level, onset, Common:Widespread) %>%
  mutate(source = 'KK10',
         level = if_else(level == 'Common',
                         'Common (> 1% land area)',
                         'Widespread (> 20% land area)')) %>%
  rename(Region = ID)
```

As above, calculate the earliest onset time for intensive agriculture at common and widespread thresholds from the consensus assessment.

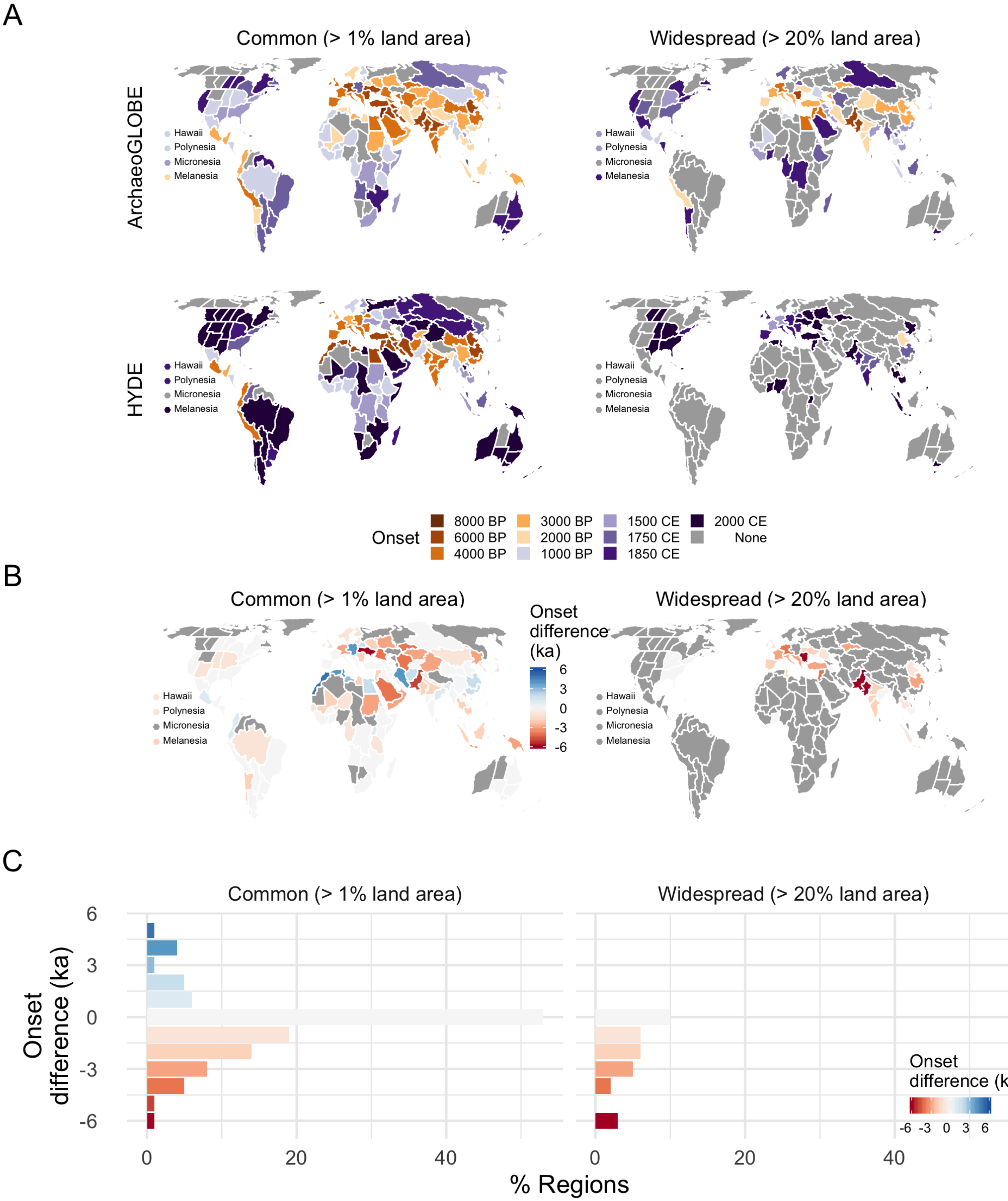
```
archaeoglobe_onset <- consensus %>%
  select(Region, Label, INAG_10KBP:INAG_1850CE) %>%
  gather(time_step, value, INAG_10KBP:INAG_1850CE) %>%
  filter(value %in% c('Common', 'Widespread')) %>%
  mutate(time = parse_number(time_step),
         time = case_when(time == 1500 ~ .45,
                          time == 1750 ~ .2,
                          time == 1850 ~ .1,
                          time <= 10 ~ time),
         time = time * -1) %>%
  group_by(Region, value) %>%
  summarise(onset = min(time)) %>%
  spread(value, onset) %>%
  mutate(Common = if_else(is.na(Common) | (Common > Widespread & !is.na(Widespread)), Widespread, Common
  gather(level, onset, Common:Widespread) %>%
  mutate(source = 'ArchaeoGLOBE',
         level = if_else(level == 'Common',
                         'Common (> 1% land area)',
                         'Widespread (> 20% land area)'))
```

Combine the HYDE and ArchaeoGLOBE onsets into a single data frame for plotting. Limit the analysis to regions that have crops in HYDE at 2000CE. Then calculate the differences in onset times between the two datasets and plot the results.

```
onsets <- bind_rows(hyde_onset, archaeoglobe_onset)

# world regions that have crops in HYDE at 2000CE
hyde_ag_regions <- hyde_crop_prop %>%
  filter(time == 0.05 & prop >= 0.01) %>%
  pull(ID)

onset_difference <- onsets %>%
  spread(source, onset) %>%
  filter(Region %in% hyde_ag_regions) %>%
  mutate(diff = ArchaeoGLOBE - HYDE,
         diff = round(diff))
```


A
Common (> 1% land area)
Widespread (> 20% land area)
ArchaeoGLOBE
HYDE
Hawaii
Polynesia
Micronesia
Melanesia
Onset
8000 BP
6000 BP
4000 BP
3000 BP
2000 BP
1000 BP
1500 CE
1750 CE
1850 CE
2000 CE
None
B
Common (> 1% land area)
Widespread (> 20% land area)
Onset difference (ka)
6
3
0
-3
-6
C
Common (> 1% land area)
Widespread (> 20% land area)
Onset difference (ka)
6
3
0
-3
-6
0
20
40
% Regions
Onset difference (k
-6 -3 0 3 6

Repeat the above the analysis for the KK10 data and plot the results.

Abandonment of Foraging

Was the abandonment of widespread foraging more correlated closely with the spread of pastoralism than crop agriculture?

To investigate weather the abandonment of widespread foraging was more correlated closely with the spread of pastoralism than crop agriculture, we computed an odds ratio using the consensus responses for foraging, pastoralism and crop agriculture for all regions during the middle and late Holocene. Odds ratios are used to compare the relative odds of the occurrence of an outcome of interest (i.e spread of pastoralism), given a condition of the variable of interest (i.e. abandonment of widespread foraging) (Szumilas 2010). We created a table of counts of regions that show a decline in foraging over time, and counts of regions where pastoralism is more widespread than intensive agriculture at an arbitrary time point, such as 2 k BP. We then computed an odds ratio for this table, and if the result is greater than one, we can conclude that the outcome of pastoralism more widespread than crop agriculture after widespread foraging is abandoned is more likely that an alternative outcome.

```
consensus_cat <-
  consensus %>%
  # convert consensus variables to ordinal factors
  mutate_at(.vars = vars(FHG_10KBP:URBAN_1850CE),
            .funs = funs(case_when(. == "Widespread" ~ 3,
                                   . == "Common" ~ 2,
                                   . == "Minimal" ~  1,
                                   . == "None" ~ 0))) %>%
  mutate_at(.vars = vars(FHG_10KBP:URBAN_1850CE),
            .funs = funs(factor(., ordered = TRUE)))

# odds ratio approach
consensus_cat_df <-
consensus_cat %>%
  # label those regions that show a decline in foraging over time
  mutate(shows_decline_in_foraging =  as.numeric((FHG_10KBP > FHG_2KBP)) ) %>%
  # label those regions that show pastoralism more widespread than crop agriculture
  mutate(shows_more_pastoralism_than_crop =   as.numeric(PAS_2KBP > INAG_2KBP )) %>%
  # check
  select( shows_decline_in_foraging,
          shows_more_pastoralism_than_crop) %>%
  group_by(shows_decline_in_foraging,
           shows_more_pastoralism_than_crop) %>%
  tally() %>%
  spread(shows_more_pastoralism_than_crop, n, fill = 0) %>%
  arrange(desc(shows_decline_in_foraging)) %>%
  select(shows_decline_in_foraging, `1`, `0`)

# show a table
consensus_cat_df_show <- consensus_cat_df
names(consensus_cat_df_show) <- c(" ",
                                  "pastoralism more widespread than crops",
                                  "pastoralism less widespread than crop")
consensus_cat_df_show$` ` <- c('shows a decline in foraging over time',
                              'shows no decline in foraging over time')
knitr::kable(consensus_cat_df_show)
```

	pastoralism more widespread than crops	pastoralism less widespread than crop
shows a decline in foraging over time	29	39
shows no decline in foraging over time	18	60

```
# get odds ratio and p-value
tab <- as.matrix(consensus_cat_df[,2:3])
ft <- fisher.test(tab)

# another way
d <- data.frame(g=factor(1:2),
                s=tab[c(1,3)],
                f=tab[c(2,4)])
g <- glm(s/(s+f) ~ g,
         weights = s + f,
         data = d,
         family="binomial")
# coef(summary(g))["g2",c("Estimate","Pr(>|z|)")]
# To get the likelihood ratio test (slightly more accurate
# than the Wald  -value shown above), do
lrt <- anova(g,test="Chisq")
p_value <- round(lrt $`Pr(>Chi)`[2], 3)

# odds ratio, check it by hand
A <- tab[1]
B <- tab[3]
C <- tab[2]
D <- tab[4]
or <- (A/B) / (C/D)
```

The odds ratio for this table is 2.479, with a p-value of 0.011. This indicates that that claim of pastoralism being more widespread than crop agriculture after widespread foraging is abandoned is supported by the data.

Szumilas, Magdalena (2010). “Explaining Odds Ratios”. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*. 19 (3): 227–229