

Contents

1 Climate futures for Haskell Indian Nations University	2
1.1 Prepare your environment	2
1.2 Finding basic layers in OpenStreetMap	3
1.3 Build a basemap from OpenStreetMap	9
1.4 Mount the climate dataset	17
1.5 Find and transfer climate data from an API	21
1.6 Organize climate data	25

```
options(knitr.kable.NA = '')
```

1 Climate futures for Haskell Indian Nations University

1.1 Prepare your environment

1. Install Climate Futures Toolbox

```
install.packages("cft")
```

2. Load other packages

```
library(tidyverse)
library(tidync)
library(cft)
library(sf)
library(ggplot2)
library(ggthemes)
library(ggpattern)
library(magick)
library(future)
library(tidytable)
library(janitor)
options(timeout = 600)
library(ggfortify)
library(forecast)
library(reticulate)
library(osmdata)
library(osmextract)
```

```
conda_update(conda = "auto")
py_install("numpy")
py_install("matplotlib")
py_install("pandas")
py_install("statsmodels")
py_install("osmnx")
py_install("geopandas")
```

1.1.1 Set your color palette

1. Download the Haskell university logo

```
seamless_image_filenames <- c(  
  'Haskell_logo_copy.png'  
)
```



2. Sample the colors on that logo to make a custom color palette for our basemap

```
our_blue <- "#3A4E8B"  
our_yellow <- "#FFD60F"  
our_beige <- "#EDEDF0"  
our_purple <- "#3E1471"
```

1.2 Finding basic layers in OpenStreetMap

Explain the basics of APIs and the premise of fast downloads.

1.2.1 Use plain language to request a bounding box

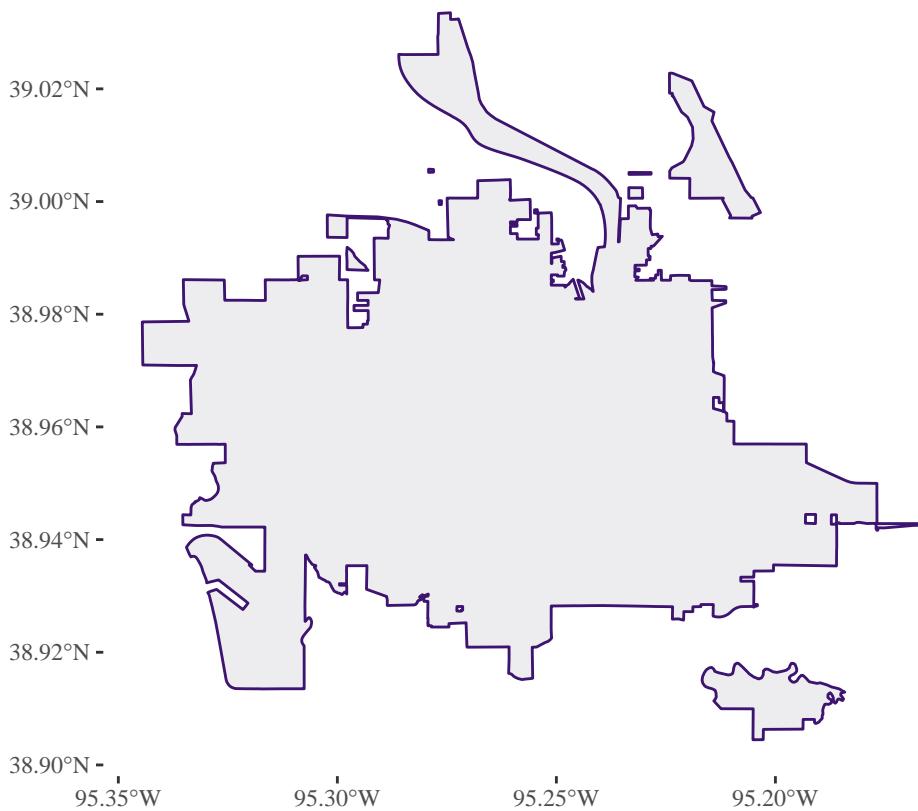
1. Find the general area on Open Street Map We use a function from the osmdata package to find a bounding box for our area of interest. This is a nice function for this purpose because it can use plain language declarations (e.g. “Lawrence, Kansas” or “Boulder, Colorado”) for the location. You do not need to use this function to define a bounding box. You can define your bounding box from any source. The benefit of this method is that it is rather easy and reliable.

```
bb <- getbb("Lawrence, Kansas")
bb
```

```
##           min         max
## x -95.34454 -95.16662
## y  38.90447  39.03350
```

If you format your request a little differently, then it will return the more complex polygon for the bounding box.

```
bb_sf <- getbb("Lawrence, Kansas", format_out = "sf_polygon")
ggplot(data=bb_sf$multipolygon) + geom_sf(fill=our_beige, color=our_purple)
  + theme_tufte()
```



2. Find any buildings associated with any University in our “Lawrence, Kansas” bounding box.

1.2.1.1 Using the package osmdata This request first calls the opq() function, which mounts to the OpenStreetMap database and then queries the “building” key (i.e. all the building footprints) for

any building types matching the value “university”. This is a representation of the “key” and “value” system that OSM uses to query data. The final step is to convert the OSM output into a spatial format that works well in R, called sf.

```
library(osmdata)
University_buildings <-
  opq(bb) %>%
  add_osm_feature(key = "building", value = "university") %>%
  osmdata_sf()
```

The output from this request shows a list of multipolygons, polygons, linestrings, and points. Each of these data types have a different storage structure, so we can't look at them all at the same time. Instead, lets start with polygons, which likely represent a single building footprint. Printing ‘my_boundary\$osm_polygons’ shows that there are two Universities in Lawrence and we need to filter those results down to only include Haskell building.

```
University_buildings <- University_buildings$osm_polygons
```

1.2.1.2 Using the package osmextract The package OSMextract is calling the same OSM API, but it does it in a slightly different way that can make it faster but also requires a little better understanding of what you're looking for. For example, by default OSM extract only downloads the first 25 columns as their own column and clumps the rest into a list that is difficult to read. You can add columns to the api call as I did here (e.g. extra_tags = c("operator")) and it will return that column as a column instead of in the list. You can also parse the list yourself or use osmdata() to download a sample and then use osmextract to execute larger downloads. This package usually clumps polygons and multipolygons by default so, you just ask for multipolygons and get both back.

```
library(osmextract)
University_buildings <- oe_get(
  place = "Lawrence, Kansas",
  layer = "multipolygons",
  query = "SELECT * FROM multipolygons WHERE building IN ('university')",
  quiet = TRUE,
  extra_tags = c("operator")
)

colnames(University_buildings)
```

```
## [1] "osm_id"      "osm_way_id"   "name"        "type"        "aeroway"
## [6] "amenity"     "admin_level"   "barrier"     "boundary"    "building"
## [11] "craft"       "geological"   "historic"    "land_area"   "landuse"
## [16] "leisure"     "man_made"     "military"    "natural"     "office"
## [21] "place"       "shop"        "sport"       "tourism"    "operator"
## [26] "other_tags"  "geometry"
```

3. Use the 'operator' column to identify the owner of those buildings and filter down to building operated by Haskell Indian Nations University.

```
Haskell_university_buildings <- University_buildings %>%
  filter(operator == "Haskell Indian Nations University")
Haskell_university_buildings1 <- Haskell_university_buildings[1,] #take the
  ↳ first building (e.g. first row) of the returns
head(Haskell_university_buildings1)
```

```
## # A tidytable: 1 x 27
##   osm_id osm_way_id name   type aeroway amenity admin~1 barrier bound~2 buil
##   <chr>  <chr>     <chr>  <chr> <chr>   <chr>   <chr>  <chr>  <chr>
## 1 <NA>   172577406 Winon~ <NA>   <NA>   <NA>   <NA>   <NA>   <NA>   univ
## # ... with 17 more variables: craft <chr>, geological <chr>, historic <chr>,
## #   land_area <chr>, landuse <chr>, leisure <chr>, man_made <chr>,
## #   military <chr>, natural <chr>, office <chr>, place <chr>, shop <chr>,
## #   sport <chr>, tourism <chr>, operator <chr>, other_tags <chr>,
## #   geometry <MULTIPOLYGON [°]>, and abbreviated variable names 1: admin_leve
## #   2: boundary, 3: building
```

The janitor package is useful for performing automated cleaning tasks on your data. Here we remove all of the columns that contain no data to make our dataframe much smaller and easier to read.

```
##   osm_way_id      name   building          operator
## 1 172577406 Winona Hall university Haskell Indian Nations University
## 
## 1 "gnis:county_id"=>"045","gnis:created"=>"12/08/2008","gnis:feature_id"=>"25
## 
##                                     geometry
## 1 MULTIPOLYGON (((-95.23447 3...
```

Sensing the Earth: CFT tutorial

4. Plot our discovered footprint to visually confirm It looks like we found Winona Hall in the OpenStreetMap database. This is how we plot the perimeters associated with it.



```
basemap <- ggplot(data = st_as_sf(boundaries1)) +  
  geom_sf(fill = our_purple, color=our_yellow) +  
  geom_sf_text(aes(label = name), size=10, color=our_yellow) +  
  theme_tufte()
```

basemap



OSM in python using osmnx
The python interface for OSM uses a slightly different syntax to write the requests, but it's calling the same OSM api that the R packages call. You still submit a plain-language area-of-interest, a value, and a key. Those three inputs will return a list of points, lines, and polygons for you to use and manipulate.

```
place_name = "Lawrence, Kansas"  
  
# import osmnx  
import osmnx as ox
```

```
import geopandas as gpd
import matplotlib.pyplot as plt

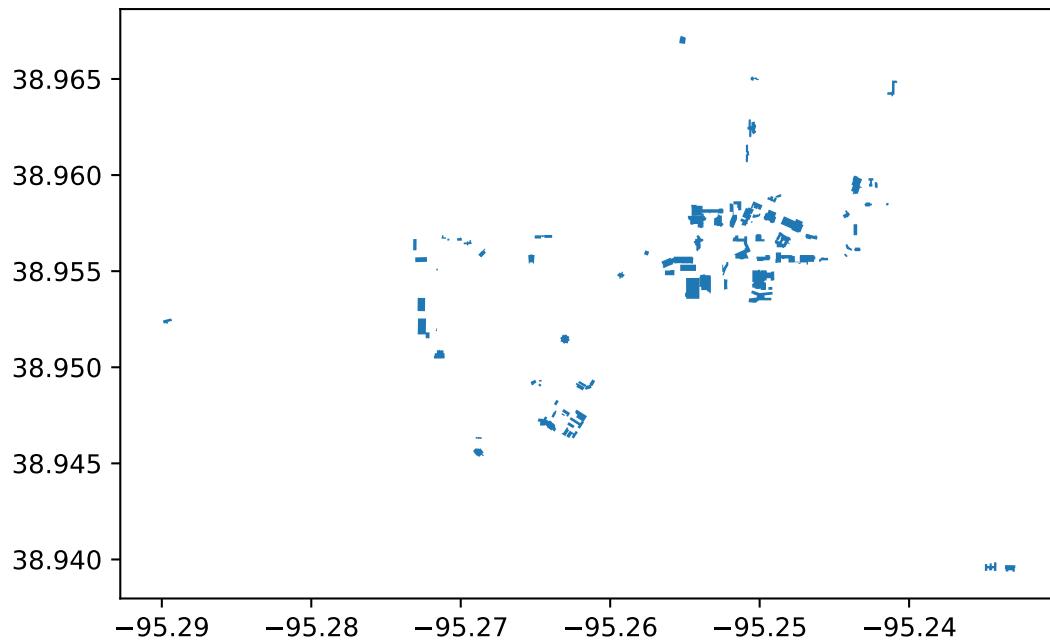
# Get place boundary related to the place name as a geodataframe
area = ox.geocode_to_gdf(place_name)
type(area)

# List key-value pairs for tags

## <class 'geopandas.geodataframe.GeoDataFrame'>

tags = {'building': 'university'}

buildings = ox.geometries_from_place(place_name, tags)
buildings.plot()
plt.show()
```



1.3 Build a basemap from OpenStreetMap

1.3.1 Download all the layers you want to include in your basemap

1. Download the Haskell University footprint

```
amenity_poly <- oe_get(  
  place = "Lawrence, Kansas",  
  layer = "multipolygons",  
  query = "SELECT * FROM multipolygons WHERE amenity IN ('university')",  
  quiet = TRUE,  
  extra_tags = c("operator"))  
)
```

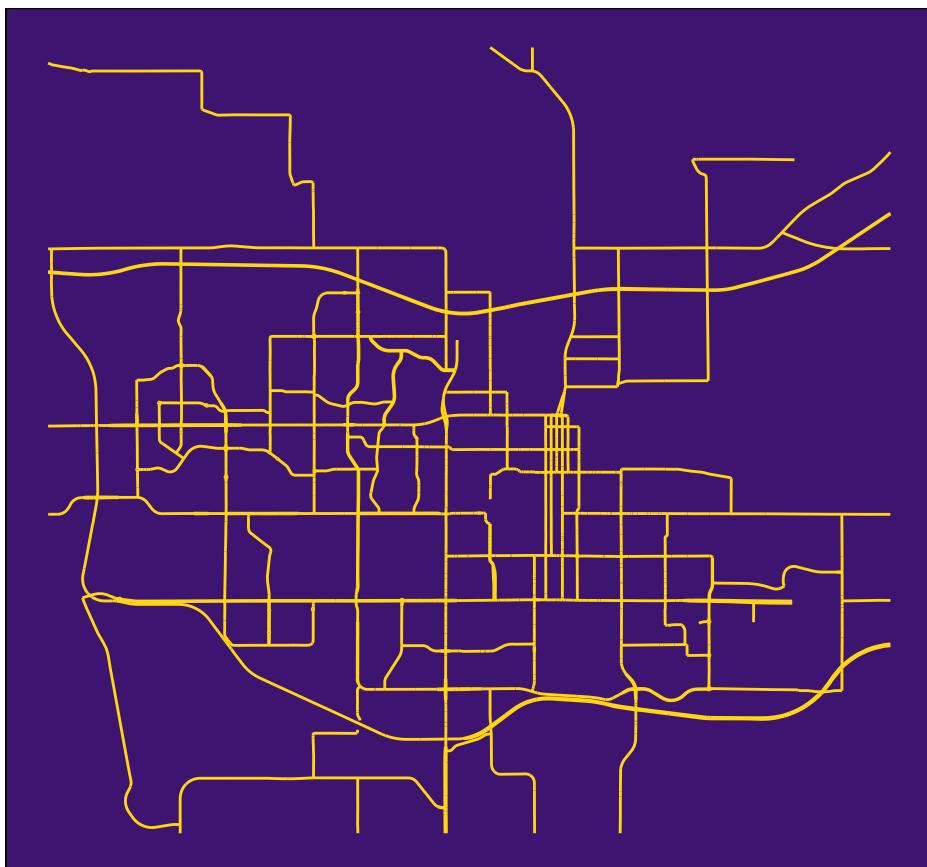
```
haskell_poly <- amenity_poly %>%  
  filter(name =='Haskell Indian Nations University') %>%  
  st_as_sf()
```



2. Download street vector layers The street vector is divided into two different downloads in order to create two different objects for coloring in the final figure. This first download will be in the foreground. It includes the larger and faster roadways.

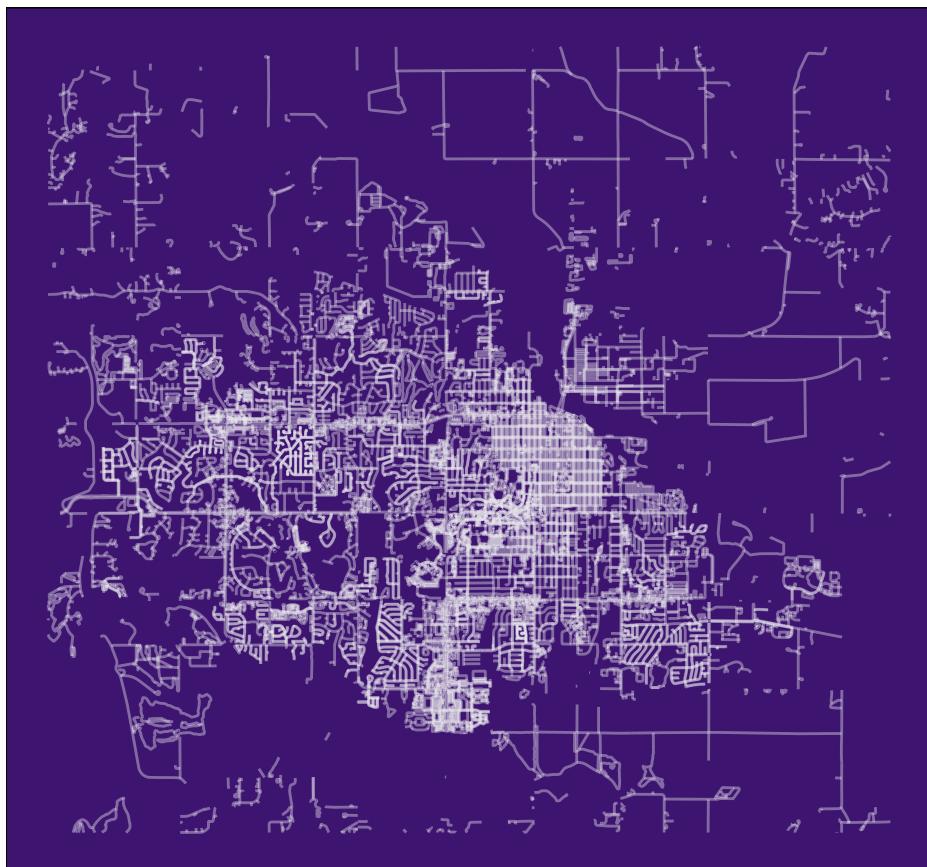
```
# the big streets
big_streets_lines <- oe_get(
  place = "Lawrence, Kansas",
  layer = "lines",
  query = "SELECT * FROM lines WHERE highway IN ('motorway', 'trunk',
    ↳ 'primary', 'secondary', 'tertiary')",
  quiet = TRUE
)

streets_crop <- big_streets_lines %>%
  st_crop(y = c(ymin = bb[2,1], ymax = bb[2,2], xmin = bb[1,1], xmax =
    ↳ bb[1,2]))
```



The second street download is for the small side streets and footpaths. These lines will be more faint and in the background.

```
small_streets <- oe_get(  
  place = "Lawrence, Kansas",  
  layer = "lines",  
  query = "SELECT * FROM lines WHERE highway IN ('residential', 'living',  
    ↪ 'unclassified', 'service', 'footway')",  
  quiet = TRUE  
)  
  
small_streets_crop <- small_streets %>%  
  st_crop(y = c(ymin = bb[2,1], ymax = bb[2,2], xmin = bb[1,1], xmax =  
    ↪ bb[1,2]))
```



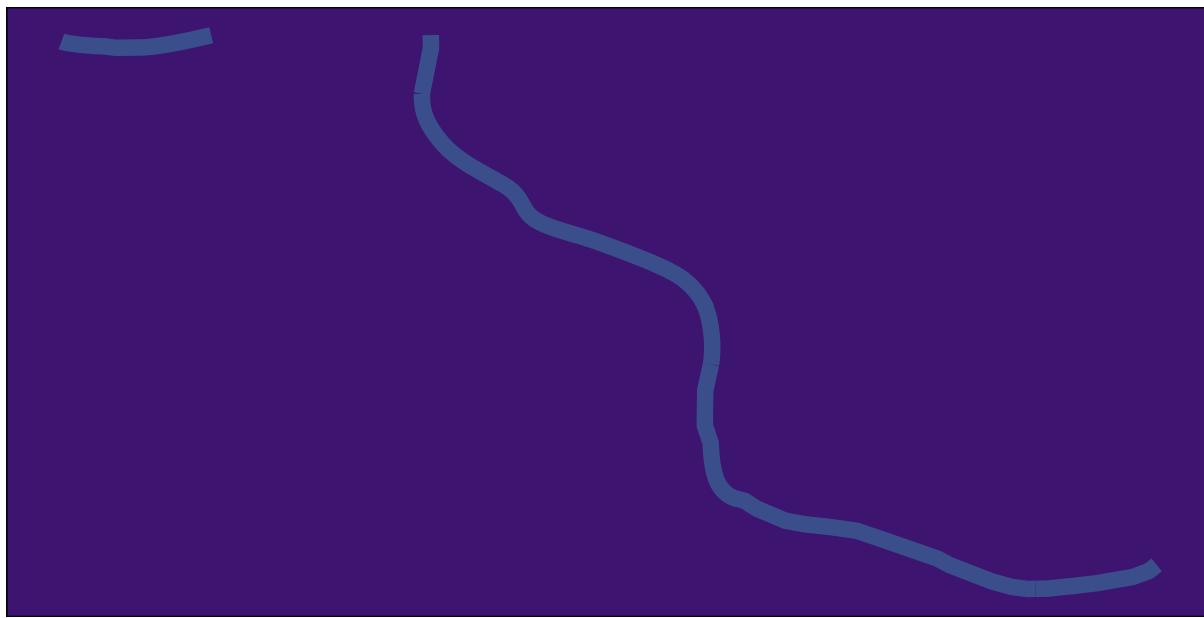
3. Download water features. The water features are first divided into moving and stationary water. We will download the river layer from the waterway key.

```
water <- oe_get(  
  place = "Lawrence, Kansas",  
  layer = "lines",
```

```
query = "SELECT * FROM lines WHERE waterway IN ('river')",
quiet = TRUE
)
```

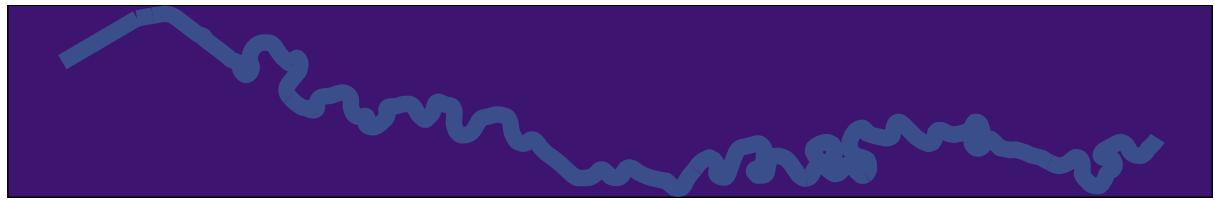
We divide the water into large and small waterways in the same way we did with the road. We are interested in making the main river much larger and the remaining waterways collectively smaller. The Kansas river is the large feature in this map so, we pull it out first.

```
Kansas_river_multi <- water %>%
  filter(name == "Kansas River") %>%
  st_as_sf() %>%
  st_crop(y = c(ymin = bb[2,1], ymax = bb[2,2], xmin = bb[1,1], xmax =
  ↴ bb[1,2]))
```



After removing the Kansas river, we are left with a number of remaining waterways that are stored as both linestrings and multilinestrings. We need to download each of those data types individually.

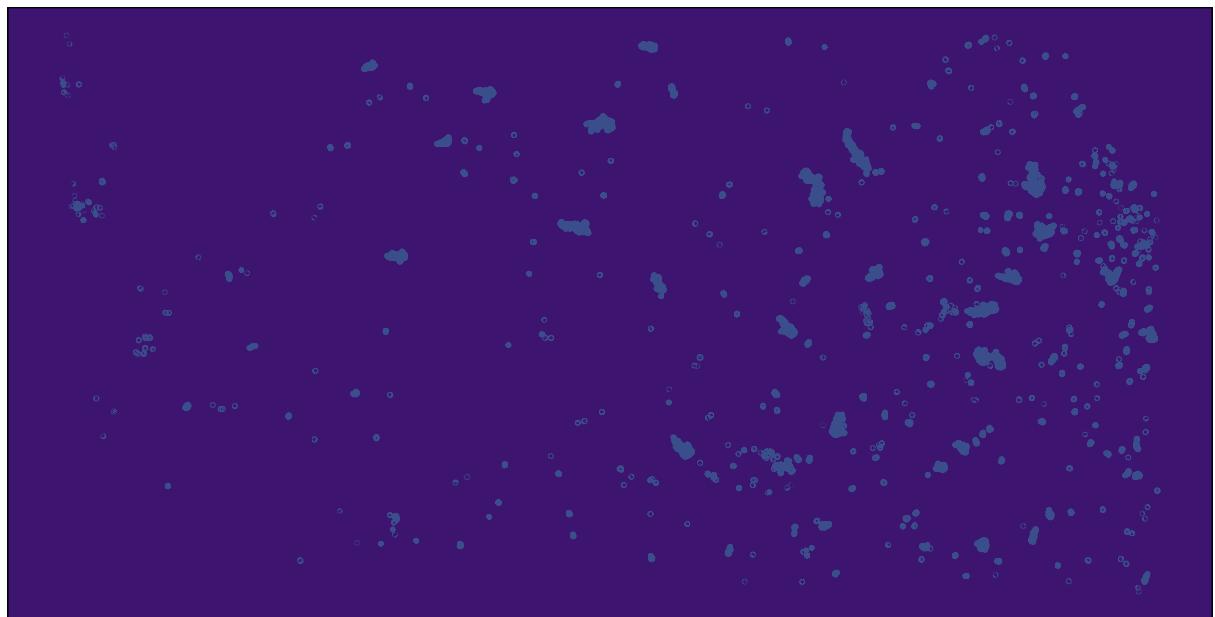
```
small_water_lines <- water %>%
  filter(name != "Kansas River")%>%
  st_as_sf() %>%
  st_crop(y = c(ymin = bb[2,1], ymax = bb[2,2], xmin = bb[1,1], xmax =
  ↴ bb[1,2]))
```



The stationary water bodies are a subcategory under the key=natural and the value=water. We ask for the extra column named water to be include in our returned sf table. We can use that column to filter our the lakes and reservours as local water bodies.

```
# Request all water features using natural:water but also request the water
# tag be given it's own column.
water_body <- oe_get(
  place = "Lawrence, Kansas",
  layer = "multipolygons",
  query = "SELECT * FROM multipolygons WHERE natural IN ('water')",
  quiet = TRUE,
  extra_tags = c("water") #give water it's own column instead of clumping
#   in supplementary list
)

water_body_crop <- water_body %>%
  filter(water == 'lake' | water == "reservoir") %>%
  st_as_sf()
```



1.3.2 Stack downloaded OSM layers into a final basemap.

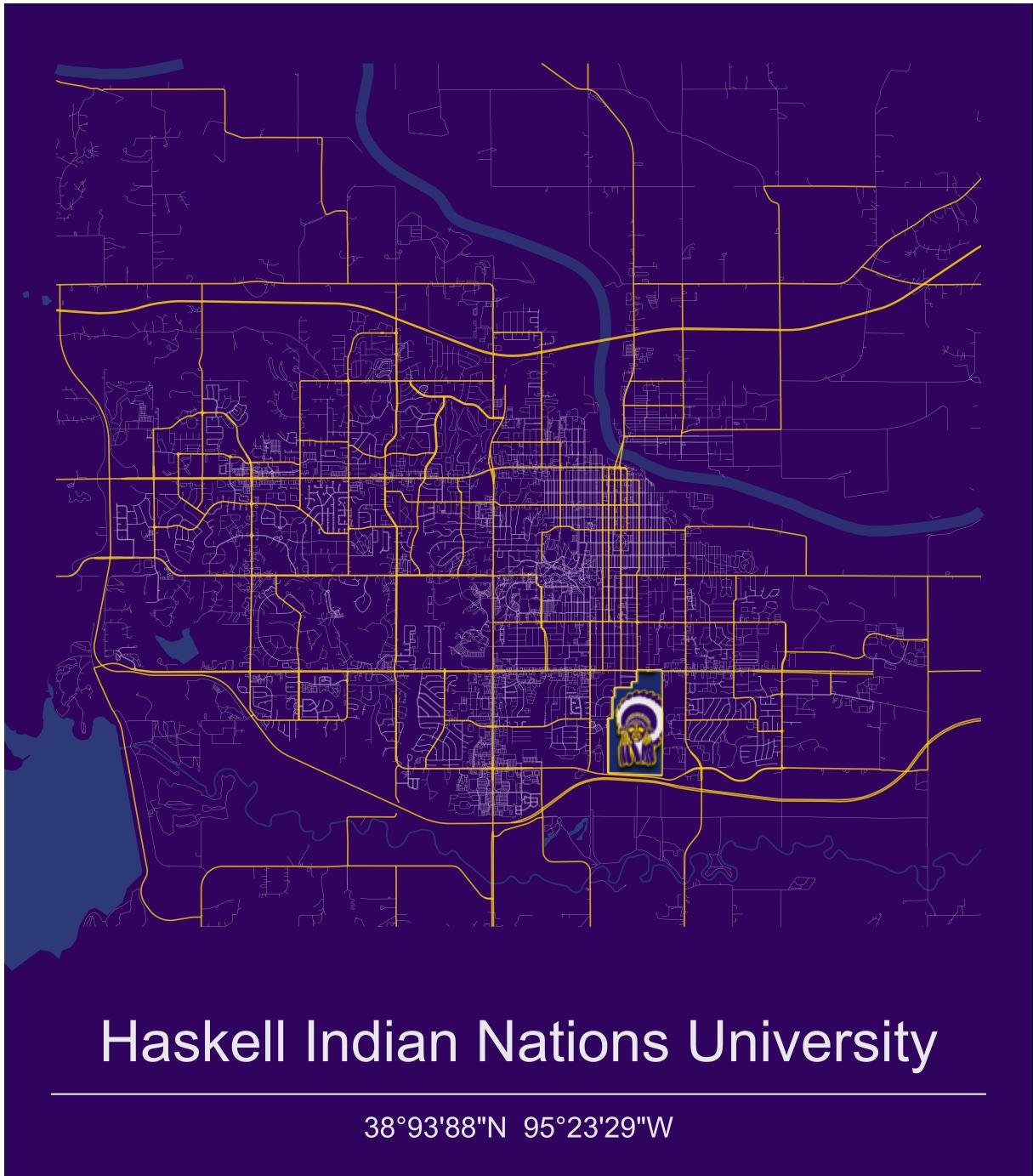
This is a special edit to manually shift the bounding box so that it better centered Haskell University in the basemap. Most people will not need this adjustment but may enjoy the ability to microadjust their basemap.

```
bbb <- bb
bbb[1,1] <- bbb[1,1] - 0.001
bbb[1,2] <- bbb[1,2] + 0.001
bbb[2,1] <- bbb[2,1] - 0.03
bbb[2,2] <- bbb[2,2] + 0.001
xlimit <- bbb[1,]
ylimit <- bbb[2,]
xmid <- xlimit[1] + diff(xlimit) / 2
ratio <- diff(xlimit) / diff(ylimit)
```

This is a long plot that calls each of the plot layers in order from the back to the front. There is a section at the end that crop, format, and append the basemap.

```
haskell_basemap <- ggplot() +
  # plot moving water layers first
  geom_sf(data = Kansas_river_multi, alpha = .8,
           size = 3, colour = our_blue) +
  geom_sf(data = small_water_lines, alpha = .8,
           size = 0.5, colour = our_blue) +
  # Layer bodies of water over the moving water layers
  geom_sf(data = water_body_crop, alpha = 1, fill = our_blue, size=0) +
  # Plot small streets in the background with a light fade
  geom_sf(data = small_streets_crop, alpha = .6,
           size = .1, colour = our_beige) +
  # Layer large streets over the top of the small streets with a bold
  # color.
  geom_sf(data = streets_crop, alpha = .8,
           size = .4, colour = our_yellow ) +
  # Layer Haskell university property polygon in the foreground
  geom_sf( data=haskell_poly, color=our_yellow, size=1) +
  # Fill Haskell property polygon with Haskell logo
  geom_sf_pattern(
    data = haskell_poly,
    size=0,
```

```
pattern      = 'image',
pattern_type = 'tile',
pattern_scale = 0.06,
pattern_filename = seamless_image_filenames
) +
# set limits on final figure
coord_sf(ylim = ylim, xlim = xlim, expand = TRUE) +
# adding labels
annotate(geom = "text", y = bbb[2,1]+ 0.013, x = xmids,
         label = "Haskell Indian Nations University", size = 12, colour =
         ↪ our_beige
) +
annotate(geom = "errorbarh", xmin = xlim[1], xmax = xlim[2], y =
         ↪ bbb[2,1]+ 0.005,
         height = 0, size = 0.5, colour = our_beige) +
annotate(geom = "text", y = bbb[2,1]+ 0.0001, x = xmids,
         label = "38°93'88\"N 95°23'29\"W", size = 6,
         colour = our_beige) +
# clean out unused elements and set background color
theme_void() +
theme(panel.background = element_rect(fill = our_purple),
      plot.background = element_rect(fill = NA))
```

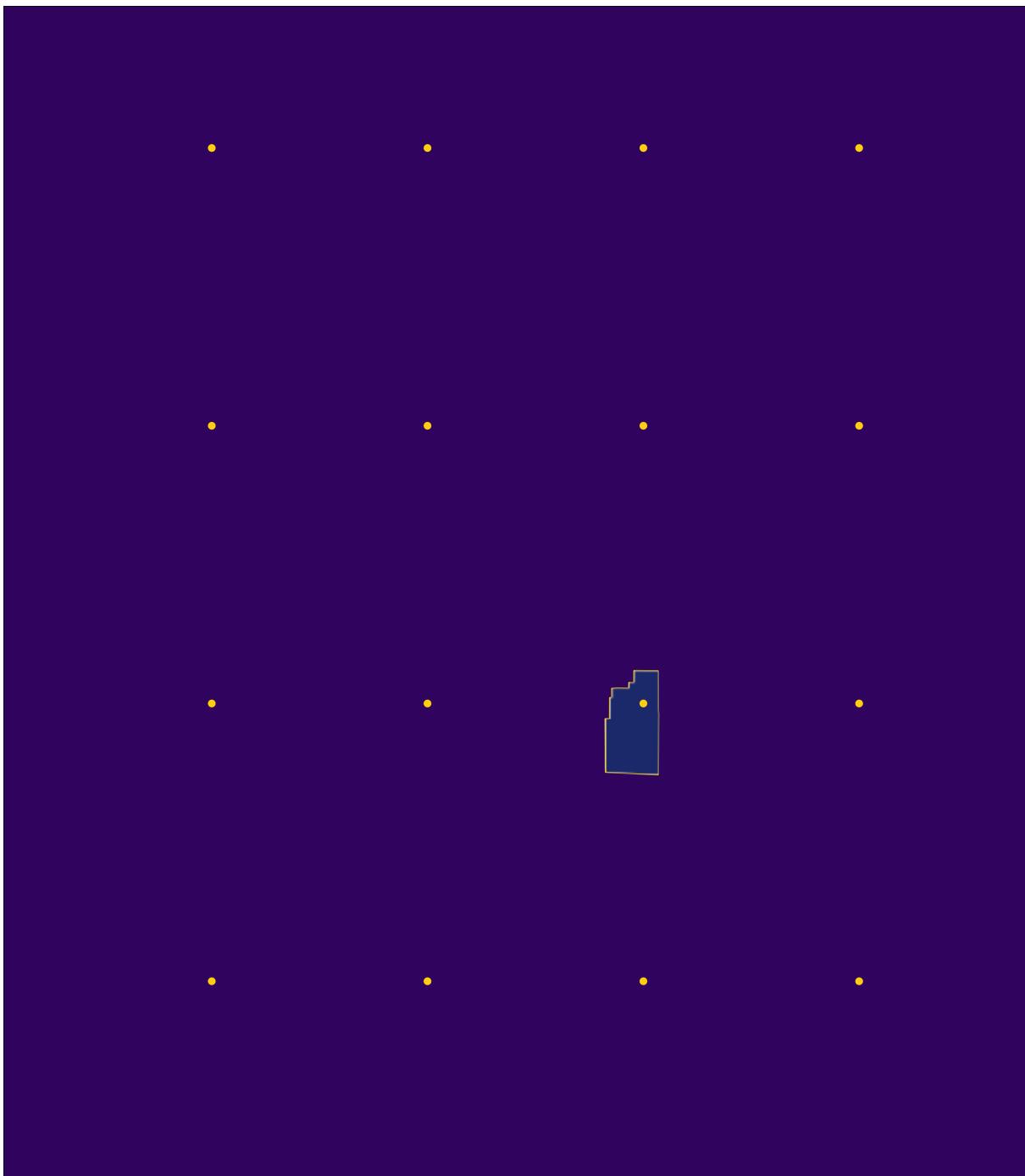


1.3.3 Save the basemap in high resolution print

```
ggsave(haskell_basemap, filename = "All_roads_lead_to_Haskell.png", height =  
↪ 11, width=8.5,  
      units="in", dpi=600)
```

1.4 Mount the climate dataset

This dataset is way too big to download to a particular machine. Instead, you mount to the analysis ready data cube (i.e. netCDF) and only download the subsetted data that you want to pull.



1. We calculate the center point for measuring to the nearest climate data point.

```
haskell_centroid <- st_coordinates(st_centroid(haskell_poly))
lat_pt <- haskel_centroid[1,2]
lon_pt <- haskel_centroid[1,1]
```

2. Connect to the web server and activate the proper data dimensions.

```
web_link = "https://cida.usgs.gov/thredds/dodsC/macav2metdata_daily_future"

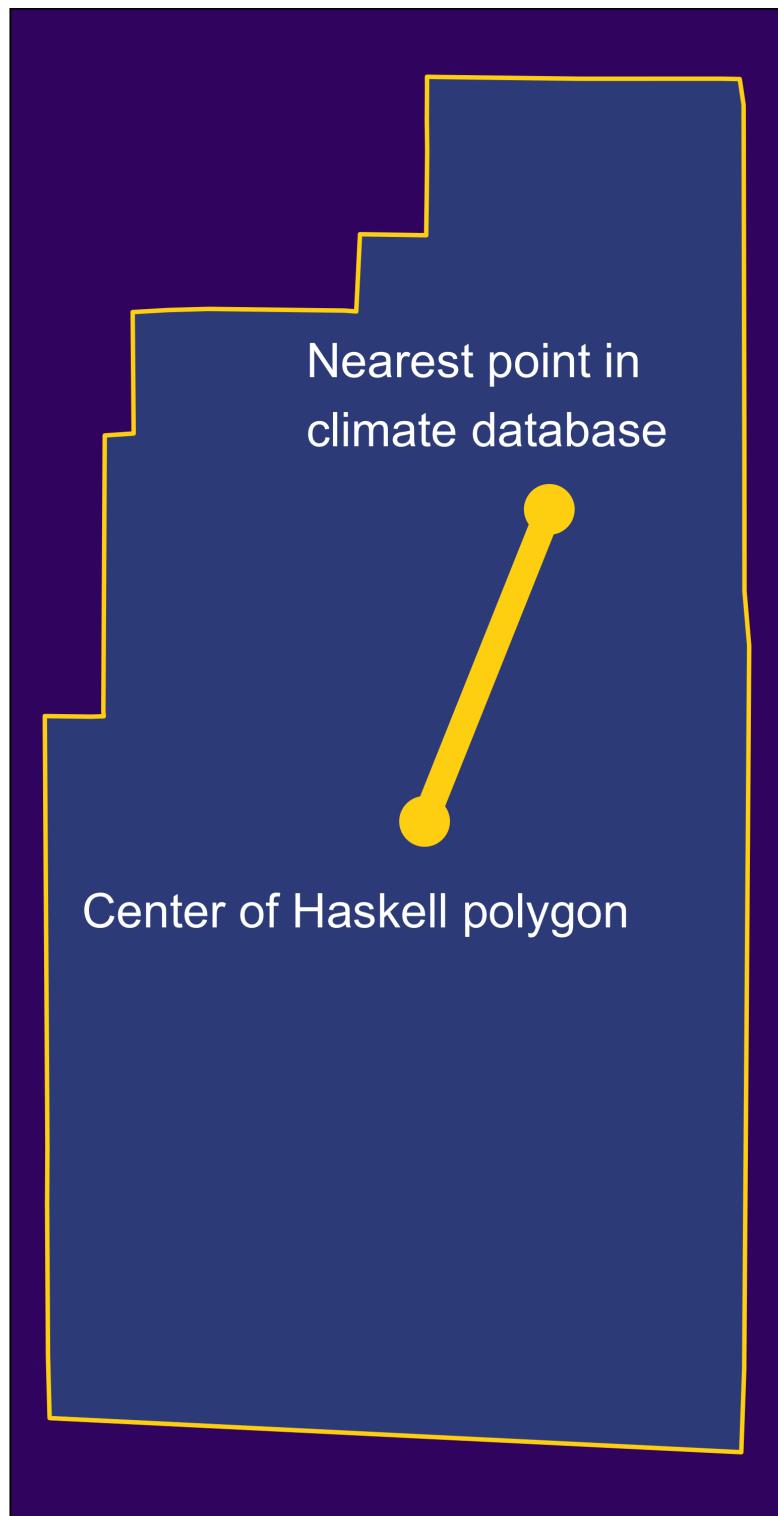
# Change to
↪ "https://cida.usgs.gov/thredds/catalog.html?dataset=cida.usgs.gov/macav2metdata_dail
↪ for historical data.

src <- tidyrc::tidyrc(web_link)
lons <- src %>% activate("D2") %>% hyper_tibble()
lats <- src %>% activate("D1") %>% hyper_tibble()
```

3. Search through the database of climate prediction points to find which one is closest to our centroid. We then spatially project that chosen pt into an sf object.

```
known_lon <- lons[which(abs(lons-lon_pt)==min(abs(lons-lon_pt))),]
known_lat <- lats[which(abs(lats-lat_pt)==min(abs(lats-lat_pt))),]

chosen_pt <- st_as_sf(cbind(known_lon,known_lat), coords = c("lon", "lat"),
↪ crs = "WGS84", agr = "constant")
```



1.5 Find and transfer climate data from an API

1.5.1 Mount to the downscaled dataset and transfer metadata.

We cannot download the entire dataset. Instead, we mount to that dataset by connecting to an api that route us to a particular part of the dataset based on the bounding box specified. Mount to the USGS downscaled dataset. The resultant object called ‘input’ includes three elements. The first is the full list of available data at each timestep. The second is a list of possible time steps. The third is a list of verbatim copy of the raw return from the server. This raw return shows the dimensions of the data and how many elements are available in each of those dimensions.

```
# Mount to the USGS downscaled dataset.  
inputs <- cft::available_data()  
  
## Trying to connect to the USGS.gov API  
  
## not a file:  
## ' https://cida.usgs.gov/thredds/dodsC/macav2metdata_daily_future '  
##  
## ... attempting remote connection  
  
## Connection succeeded.  
  
## Reading results  
  
## Converting into an R data.table
```

Element 1. Dataframe of available data and descriptions of each of those variables.

```
head(inputs[[1]])  
  
## # A tibble: 6 x 9  
##   Available variab~1 Varia~2 Units Model Model~3 Scena~4 Varia~5 Model~6 Scen  
##   <chr>          <chr>    <chr> <chr> <chr>    <chr>    <chr>    <chr>    <chr>  
## 1 huss_BNU-ESM_r1i1~ Specif~ kg k~ Beij~ r1i1p1 RCP 4.5 huss      BNU-  
ESM rcp45  
## 2 huss_BNU-ESM_r1i1~ Specif~ kg k~ Beij~ r1i1p1 RCP 8.5 huss      BNU-  
ESM rcp85
```

```
## 3 huss_CCSM4_r6i1p1~ Specif~ kg k~ Comm~ r6i1p1 RCP 4.5 huss CCSM4 rcp4
## 4 huss_CCSM4_r6i1p1~ Specif~ kg k~ Comm~ r6i1p1 RCP 8.5 huss CCSM4 rcp8
## 5 huss_CNRM-CM5_r1i~ Specif~ kg k~ Cent~ r1i1p1 RCP 4.5 huss CNRM-
C~ rcp45
## 6 huss_CNRM-CM5_r1i~ Specif~ kg k~ Cent~ r1i1p1 RCP 8.5 huss CNRM-
C~ rcp85
## # ... with abbreviated variable names 1: `Available variable`, 2: Variable,
## # 3: `Model ensemble type (only CCSM4 relevant)`, 4: Scenario,
## # 5: `Variable abbreviation`, 6: `Model abbreviation`,
## # 7: `Scenario abbreviation`
```

Element 2. Short list of available data you can request

```
head(unique(inputs[[1]]$Variable))
```

```
## [1] "Specific Humidity"                  "Precipitation"
## [3] "Maximum Relative Humidity"          "Minimum Relative Humidity"
## [5] "Surface Downswelling Shortwave Flux" "Maximum Temperature"
```

Element 3. Dataframe of available times

```
head(inputs[[2]])
```

```
##   Available times      dates
## 1             38716 2006-01-01
## 2             38717 2006-01-02
## 3             38718 2006-01-03
## 4             38719 2006-01-04
## 5             38720 2006-01-05
## 6             38721 2006-01-06
```

```
tail(inputs[[2]])
```

```
##       Available times      dates
## 34328            73043 2099-12-26
## 34329            73044 2099-12-27
```

```
## 34330      73045 2099-12-28
## 34331      73046 2099-12-29
## 34332      73047 2099-12-30
## 34333      73048 2099-12-31
```

1.5.2 Decide which Variables, Scenarios, and Models you want to request

Your data order needs to include three things: Variable, Scenario, and Model. Your options can be found in the input elements we explored in the previous section.

```
input_variables <- inputs$variable_names %>%
  filter(Variable %in% c("Maximum Relative Humidity",
                        "Minimum Relative Humidity",
                        "Maximum Temperature",
                        "Minimum Temperature",
                        "Precipitation",
                        "Eastward Wind",
                        "Northward Wind")) %>%
  filter(Scenario %in% c( "RCP 8.5")) %>%
  filter(Model %in% c(
    "Beijing Climate Center - Climate System Model 1.1",
    "Beijing Normal University - Earth System Model",
    "Canadian Earth System Model 2",
    ↪
    "Centre National de Recherches Météorologiques - Climate Model 5",
    ↪
    "Commonwealth Scientific and Industrial Research Organisation - Mk3.6.0",
    ↪
    "Community Climate System Model 4",
    ↪
    "Geophysical Fluid Dynamics Laboratory - Earth System Model 2 Generalized
     ↪ Ocean Layer Dynamics",
    "Geophysical Fluid Dynamics Laboratory - Earth System Model 2 Modular
     ↪ Ocean",
    "Hadley Global Environment Model 2 - Climate Chemistry 365 (day) ",
    ↪
    "Hadley Global Environment Model 2 - Earth System 365 (day)",
    ↪
    "Institut Pierre Simon Laplace (IPSL) - Climate Model 5A - Low
     ↪ Resolution",
    "Institut Pierre Simon Laplace (IPSL) - Climate Model 5A - Medium
     ↪ Resolution",
```

```
"Institut Pierre Simon Laplace (IPSL) - Climate Model 5B - Low
→ Resolution",
"Institute of Numerical Mathematics Climate Model 4",
→
"Metereological Research Institute - Coupled Global Climate Model 3",
→
"Model for Interdisciplinary Research On Climate - Earth System Model",
→
"Model for Interdisciplinary Research On Climate - Earth System Model -
→ Chemistry",
"Model for Interdisciplinary Research On Climate 5",
→
"Norwegian Earth System Model 1 - Medium Resolution" )) %>%
  pull("Available variable")

head(as.data.frame(input_variables))

##           input_variables
## 1      pr_BNU-ESM_r1i1p1_rcp85
## 2      pr_CCSM4_r6i1p1_rcp85
## 3      pr_CNRM-CM5_r1i1p1_rcp85
## 4 prCSIRO-Mk3-6-0_r1i1p1_rcp85
## 5      pr_CanESM2_r1i1p1_rcp85
## 6      pr_GFDL-ESM2G_r1i1p1_rcp85
```

1.5.3 Prepare for parallelization

We have requested enough information to exceed our download limit and we need to implement a ‘parallel’ approach to get all the data we want. This strategy has each of the cores in your computer act as their own computer and individually make small requests from the server and then assemble all the little chunks into the final dataset you requested. Here we tell our computer that we are about to send a parallel request and tell the computer how we want to distribute the tasks we send.

```
# ask how many cores are available to be farmed out. I subtract one so I
→ still have a core to use for controlling the whole process.
n_cores <- availableCores() - 1
```

```
# set plan to take all cores except one.  
plan(multisession, workers = n_cores)
```

1.5.4 Make parallel call to USGS server.

This is the parallel function from the CFT package. It will shuttle all the data you requested, through all the cores you specified, for the latitude and longitude you requested. This took about 45 minutes on my home laptop with fiber internet. Virtual machines usually only have one core. If you're running this in the cloud, you may need to do some special configuration to get this to actually parallelize.

```
out <- single_point_firehose(input_variables, known_lat, known_lon )  
head(out)
```

1.5.5 Save output

To save time, I have run the api request above and saved the results for future use. There is no reason to run the api request over and over again. It's easy to save the data once their downloaded and save yourself the download again in the future.

```
haskell <- out  
save(haskell, file = "haskell.RData")
```

1.5.6 Load saved output

If I have run my api request and saved it to my working directory, then I can load it from here anytime I need. This will save you time as you experiment with different downstream analyses.

```
load(file = "haskell.RData")
```

1.6 Organize climate data

1.6.1 Join data with dates

```

# make the time output into it's own dataframe and change column name
available_times <- inputs[[2]]
colnames(available_times)[1] <- "time"

# left join the time data into the spatial data
haskell_posix <- haskell %>%
  left_join(available_times, by="time")

# convert time format into POSIX, which is a format that deals with all the
# confusion of time and date formats (e.g. time zones and translation
# between numbers and word descriptions for time)
haskell_posix$dates <- as.POSIXct(haskell_posix$dates)
class(haskell_posix$dates)

## [1] "POSIXct" "POSIXt"

#reorder so that dates are the first column
haskell_posix <- haskell_posix[,c(93,2, 1,3:92)]
colnames(haskell_posix)

##   [1] "dates"                               "time"
##   [3] "pr_BNU-ESM_r1i1p1_rcp85"           "pr_CCSM4_r6i1p1_rcp85"
##   [5] "pr_CNRM-CM5_r1i1p1_rcp85"          "pr_CSIRO-Mk3-6-0_r1i1p1_rcp85"
##   [7] "pr_CanESM2_r1i1p1_rcp85"            "pr_GFDL-ESM2G_r1i1p1_rcp85"
##   [9] "pr_GFDL-ESM2M_r1i1p1_rcp85"         "pr_HadGEM2-CC365_r1i1p1_rcp85"
##  [11] "pr_HadGEM2-ES365_r1i1p1_rcp85"       "pr_IPSL-CM5A-LR_r1i1p1_rcp85"
##  [13] "pr_IPSL-CM5A-MR_r1i1p1_rcp85"        "pr_IPSL-CM5B-LR_r1i1p1_rcp85"
##  [15] "pr_MIROC-ESM-CHEM_r1i1p1_rcp85"      "pr_MIROC-ESM_r1i1p1_rcp85"
##  [17] "pr_MIROC5_r1i1p1_rcp85"                "pr_MRI-CGCM3_r1i1p1_rcp85"
##  [19] "pr_NorESM1-M_r1i1p1_rcp85"              "pr_inmcm4_r1i1p1_rcp85"
##  [21] "rhsmax_BNU-ESM_r1i1p1_rcp85"            "rhsmax_CNRM-CM5_r1i1p1_rcp85"
##  [23] "rhsmax_CSIRO-Mk3-6-0_r1i1p1_rcp85"      "rhsmax_CanESM2_r1i1p1_rcp85"
##  [25] "rhsmax_GFDL-ESM2G_r1i1p1_rcp85"          "rhsmax_HadGEM2-CC365_r1i1p1_rcp85"
##  [27] "rhsmax_HadGEM2-ES365_r1i1p1_rcp85"        "rhsmax_IPSL-CM5A-
LR_r1i1p1_rcp85"
##  [29] "rhsmax_IPSL-CM5A-MR_r1i1p1_rcp85"        "rhsmax_IPSL-CM5B-
LR_r1i1p1_rcp85"
##  [31] "rhsmax_MIROC-ESM-CHEM_r1i1p1_rcp85"      "rhsmax_MIROC-ESM_r1i1p1_rcp85"

```

```
## [33] "rhsmax_MIROC5_r1i1p1_rcp85"      "rhsmax_MRI-CGCM3_r1i1p1_rcp85"
## [35] "rhsmax_inmcm4_r1i1p1_rcp85"      "rhsmin_BNU-ESM_r1i1p1_rcp85"
## [37] "rhsmin_CNRM-CM5_r1i1p1_rcp85"      "rhsmin_CSIRO-Mk3-6-
0_r1i1p1_rcp85"
## [39] "rhsmin_CanESM2_r1i1p1_rcp85"      "rhsmin_GFDL-ESM2G_r1i1p1_rcp85"
## [41] "rhsmin_GFDL-ESM2M_r1i1p1_rcp85"      "rhsmin_HadGEM2-CC365_r1i1p1_rcp85"
## [43] "rhsmin_HadGEM2-ES365_r1i1p1_rcp85"      "rhsmin_IPSL-CM5A-
LR_r1i1p1_rcp85"
## [45] "rhsmin_IPSL-CM5A-MR_r1i1p1_rcp85"      "rhsmin_IPSL-CM5B-
LR_r1i1p1_rcp85"
## [47] "rhsmin_MIROC-ESM-CHEM_r1i1p1_rcp85"      "rhsmin_MIROC-ESM_r1i1p1_rcp85"
## [49] "rhsmin_MIROC5_r1i1p1_rcp85"      "rhsmin_MRI-CGCM3_r1i1p1_rcp85"
## [51] "tasmax_BNU-ESM_r1i1p1_rcp85"      "tasmax_CCSM4_r6i1p1_rcp85"
## [53] "tasmax_CNRM-CM5_r1i1p1_rcp85"      "tasmax_CSIRO-Mk3-6-
0_r1i1p1_rcp85"
## [55] "tasmax_CanESM2_r1i1p1_rcp85"      "tasmax_GFDL-ESM2G_r1i1p1_rcp85"
## [57] "tasmax_GFDL-ESM2M_r1i1p1_rcp85"      "tasmax_HadGEM2-CC365_r1i1p1_rcp85"
## [59] "tasmax_HadGEM2-ES365_r1i1p1_rcp85"      "tasmax_IPSL-CM5A-
LR_r1i1p1_rcp85"
## [61] "tasmax_IPSL-CM5A-MR_r1i1p1_rcp85"      "tasmax_IPSL-CM5B-
LR_r1i1p1_rcp85"
## [63] "tasmax_MIROC-ESM-CHEM_r1i1p1_rcp85"      "tasmax_MIROC-ESM_r1i1p1_rcp85"
## [65] "tasmax_MIROC5_r1i1p1_rcp85"      "tasmax_MRI-CGCM3_r1i1p1_rcp85"
## [67] "tasmax_NorESM1-M_r1i1p1_rcp85"      "tasmax_inmcm4_r1i1p1_rcp85"
## [69] "tasmin_BNU-ESM_r1i1p1_rcp85"      "tasmin_CCSM4_r6i1p1_rcp85"
## [71] "tasmin_CNRM-CM5_r1i1p1_rcp85"      "tasmin_CSIRO-Mk3-6-
0_r1i1p1_rcp85"
## [73] "tasmin_CanESM2_r1i1p1_rcp85"      "tasmin_GFDL-ESM2G_r1i1p1_rcp85"
## [75] "tasmin_GFDL-ESM2M_r1i1p1_rcp85"      "tasmin_HadGEM2-CC365_r1i1p1_rcp85"
## [77] "tasmin_HadGEM2-ES365_r1i1p1_rcp85"      "tasmin_IPSL-CM5A-
LR_r1i1p1_rcp85"
## [79] "tasmin_IPSL-CM5A-MR_r1i1p1_rcp85"      "tasmin_IPSL-CM5B-
LR_r1i1p1_rcp85"
## [81] "tasmin_MIROC-ESM-CHEM_r1i1p1_rcp85"      "tasmin_MIROC-ESM_r1i1p1_rcp85"
## [83] "tasmin_MIROC5_r1i1p1_rcp85"      "tasmin_MRI-CGCM3_r1i1p1_rcp85"
## [85] "tasmin_NorESM1-M_r1i1p1_rcp85"      "tasmin_inmcm4_r1i1p1_rcp85"
## [87] "pr_bcc-csm1-1_r1i1p1_rcp85"      "rhsmax_bcc-csm1-1_r1i1p1_rcp85"
## [89] "rhsmin_bcc-csm1-1_r1i1p1_rcp85"      "rhsmin_inmcm4_r1i1p1_rcp85"
```

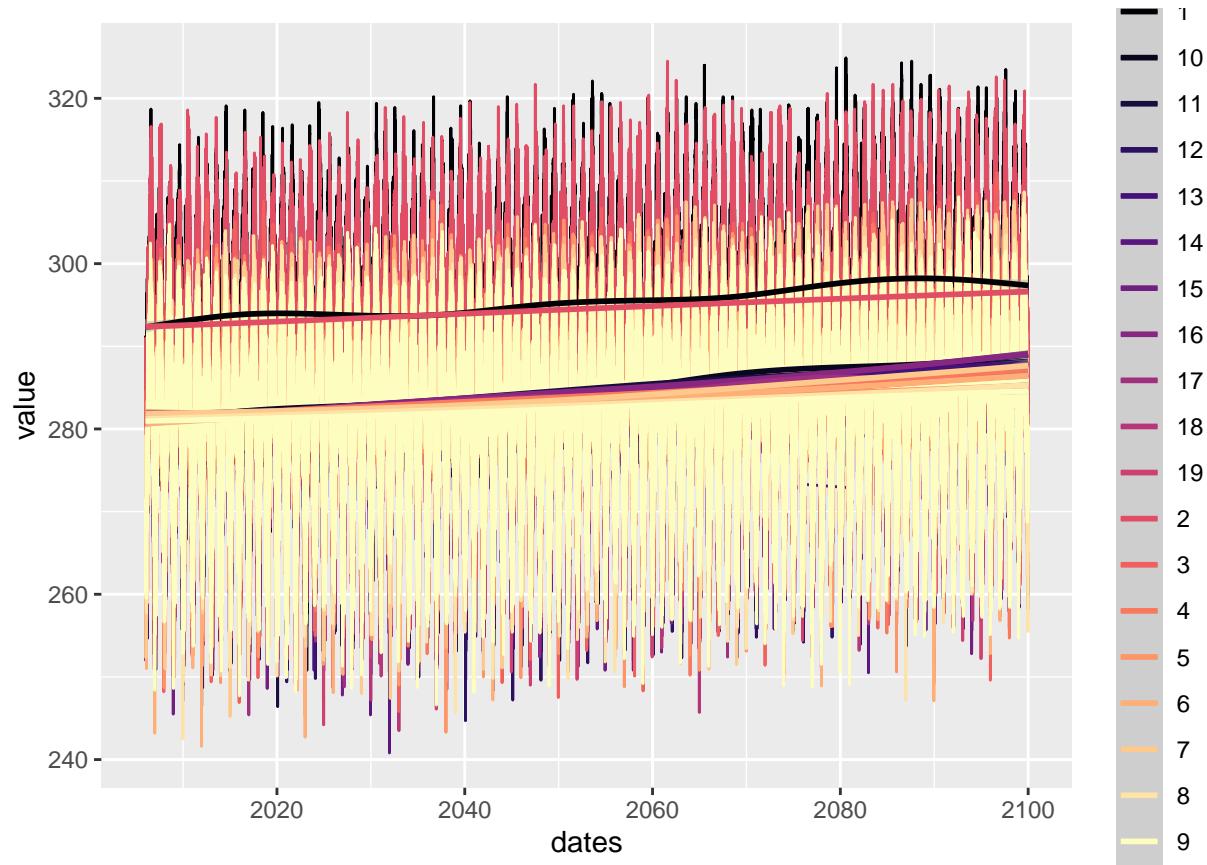
```
## [91] "tasmax_bcc-csm1-1_r1i1p1_rcp85"      "tasmin_bcc-csm1-1_r1i1p1_rcp85"  
## [93] "geometry"
```

1.6.2 Temperature

```
df_min_temp <- haskell_posix %>%  
  st_drop_geometry() %>%  
  select(dates, dates, which(startsWith(colnames(haskell), "tasmin"))) %>%  
  gather(key = "variable", value = "value", -dates)  
  
df_min_temp <- df_min_temp[which(df_min_temp$value > 200), ]  
df_min_temp$variable <-  
  ↪ as.character(as.numeric(as.factor(df_min_temp$variable)))  
  
#df_min_temp_TS <- as.xts(df_min_temp)  
head(df_min_temp)
```

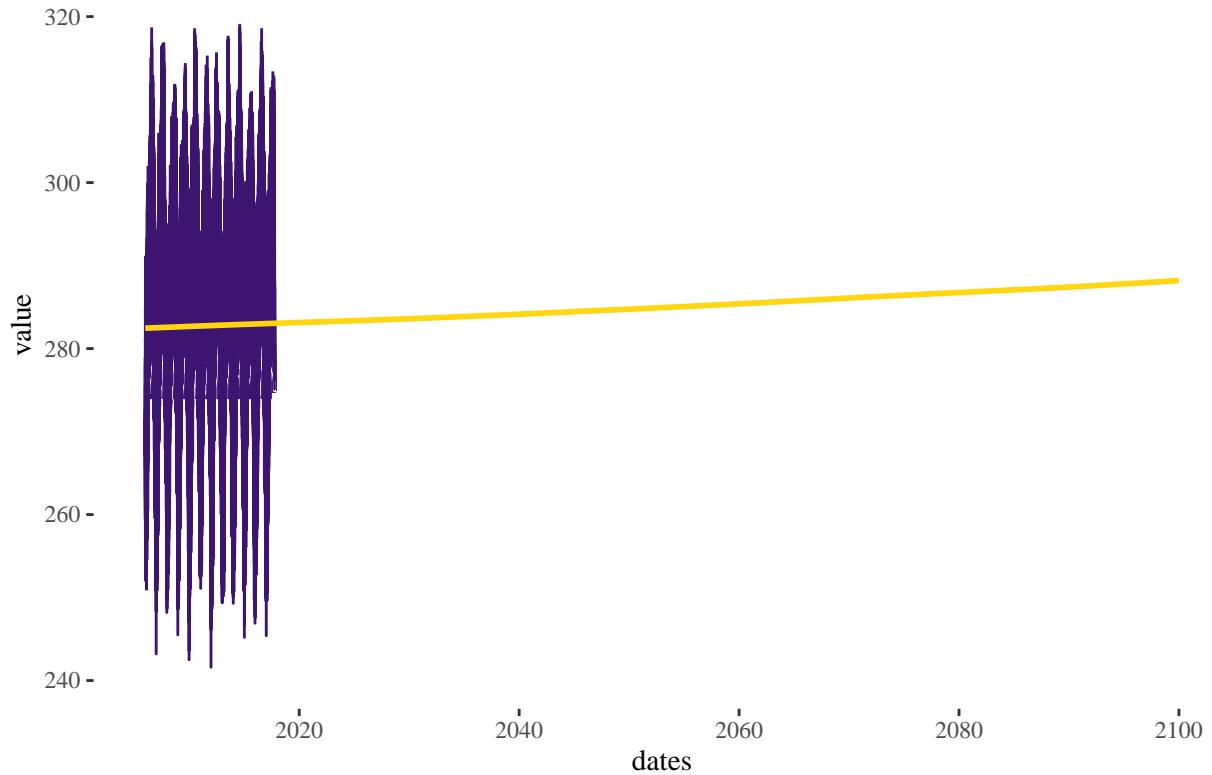
1.6.2.1 Minimum temperature

```
##           dates variable value  
## 1 2006-01-01        2 273.9  
## 2 2006-01-02        2 280.6  
## 3 2006-01-03        2 280.9  
## 4 2006-01-04        2 284.3  
## 5 2006-01-05        2 290.5  
## 6 2006-01-06        2 291.0  
  
all_climate_projections <- ggplot(data= df_min_temp, aes(x = dates, y =  
  ↪ value, color = variable)) +  
  geom_line() +  
  geom_smooth() +  
  scale_colour_viridis_d(option="A")  
  
all_climate_projections
```



```
ensemble_climate_projections <- ggplot(data= df_min_temp, aes(x = dates, y
  = value)) +
  geom_line(color=our_purple) +
  geom_smooth(color=our_yellow) +
  theme_tufte()
```

```
ensemble_climate_projections
```



```

library(stats)
convert.fft <- function(cs, sample.rate=1) {
  cs <- cs / length(cs) # normalize

  distance.center <- function(c)signif( Mod(c),      4)
  angle           <- function(c)signif( 180*Arg(c)/pi, 3)

  df <- data.frame(cycle    = 0:(length(cs)-1),
                   freq     = 0:(length(cs)-1) * sample.rate / length(cs),
                   strength = sapply(cs, distance.center),
                   delay    = sapply(cs, angle))
  df
}

fft_out <- convert.fft(fft(df_min_temp$value))
head(fft_out)

##   cycle         freq strength  delay
## 1 0 0.000000e+00 285.1000  0.000
## 2 1 1.537281e-06  1.0160 10.700

```

```
## 3      2 3.074563e-06  1.0860 -4.560
## 4      3 4.611844e-06  0.9302  2.350
## 5      4 6.149126e-06  1.1560 -1.210
## 6      5 7.686407e-06  0.6849  0.778

library(reticulate)
x <- r_to_py(df_min_temp$time)
y <- r_to_py(df_min_temp$value)
minTemp <- r_to_py(df_min_temp[,c(1,3)])

import numpy as np
import matplotlib.pyplot as plt

x = list(range(len(r.x)))
y = r.y

# apply fast fourier transform and take absolute values
f=abs(np.fft.fft(r.y))

# get the list of frequencies
num=np.size(x)
freq = [i / num for i in list(range(num))]

# get the list of spectrums
spectrum=f.real*f.real+f.imag*f.imag
nspectrum=spectrum/spectrum[0]

plt.clf()
# plot nspectrum per frequency, with a semilog scale on nspectrum
plt.semilogy(freq,nspectrum)

plt.show()

# improve the plot by adding periods in number of weeks rather than
# frequency
import pandas as pd
results = pd.DataFrame({'freq': freq, 'nspectrum': nspectrum})
results['period'] = results['freq'] / (1/52)
```

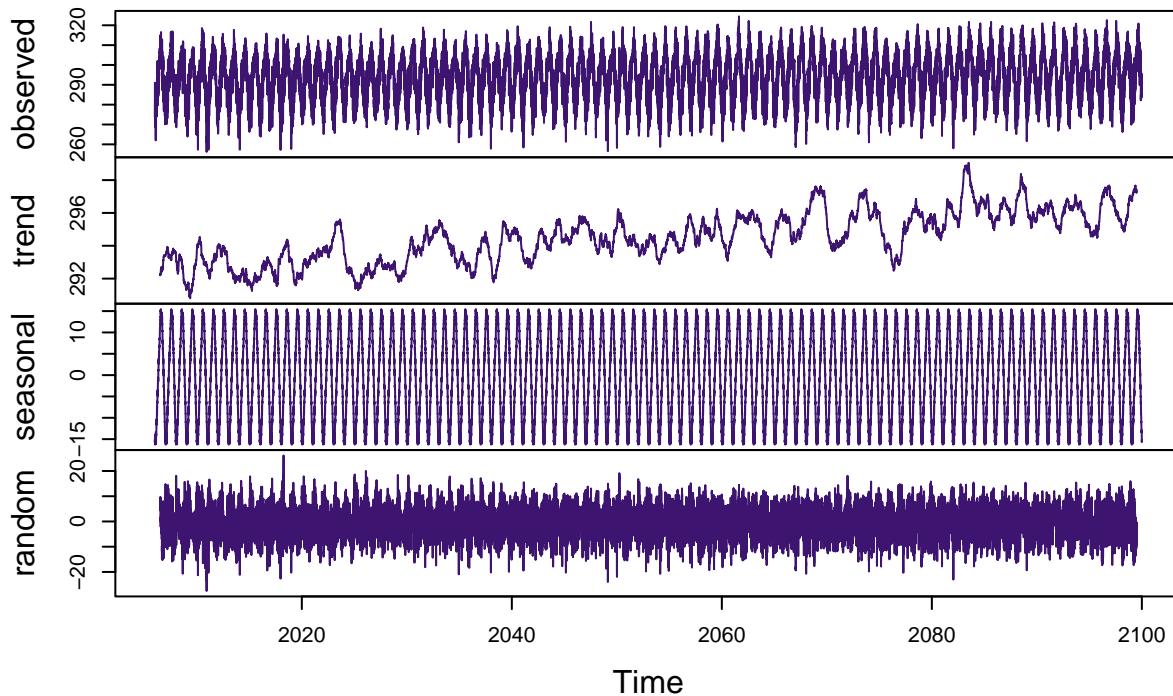
```
plt.clf()
plt.semilogy(results['period'], results['nspectrum'])
plt.show()

# improve the plot by converting the data into grouped per week to avoid
# peaks
results['period_round'] = results['period'].round()
grouped_week = results.groupby('period_round')['nspectrum'].sum()

plt.clf()
plt.semilogy(grouped_week.index, grouped_week)
plt.xticks([1, 13, 26, 39, 52])
plt.show()

temp <- ts(df_min_temp[,3], frequency = 365, start = c(2006, 1), end =
# c(2099,365))
decomp_temp <- decompose(temp)
plot(decomp_temp, col=our_purple)
```

Decomposition of additive time series

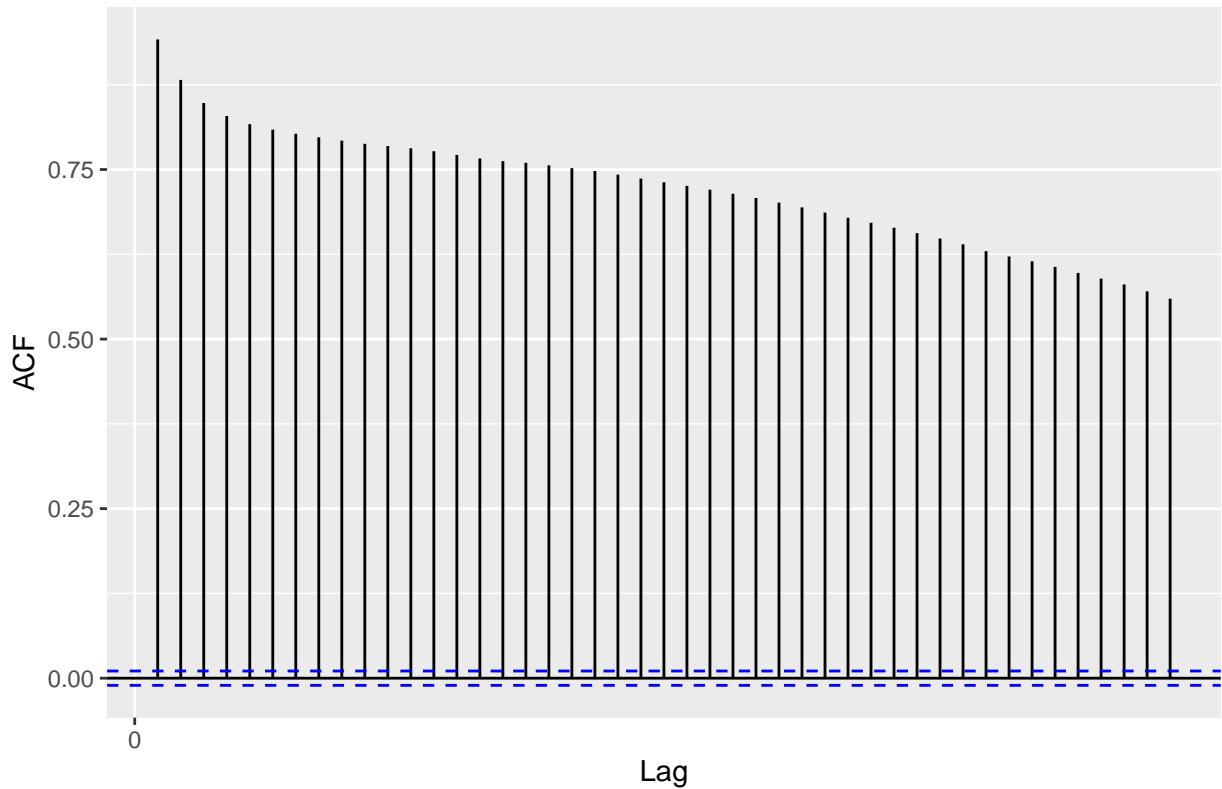


```
ggplot(akl_daily, aes(x = max_temp, y = month, height = ..density..)) +  
  geom_density_ridges(stat = "density")
```

```
library(ggfortify)  
autoplot(stl(temp, s.window = 'periodic'), ts.colour = our_purple)
```

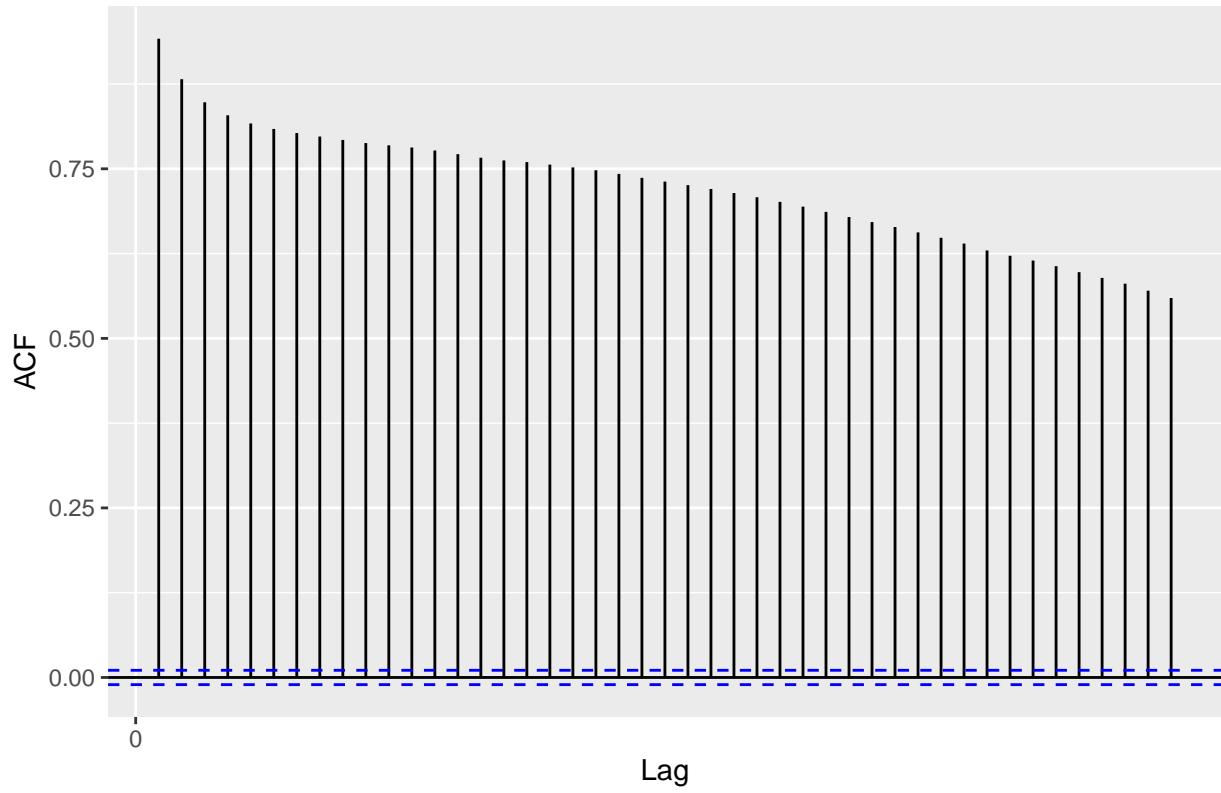
```
autoplot(acf(temp, plot = FALSE))
```

Series: temp

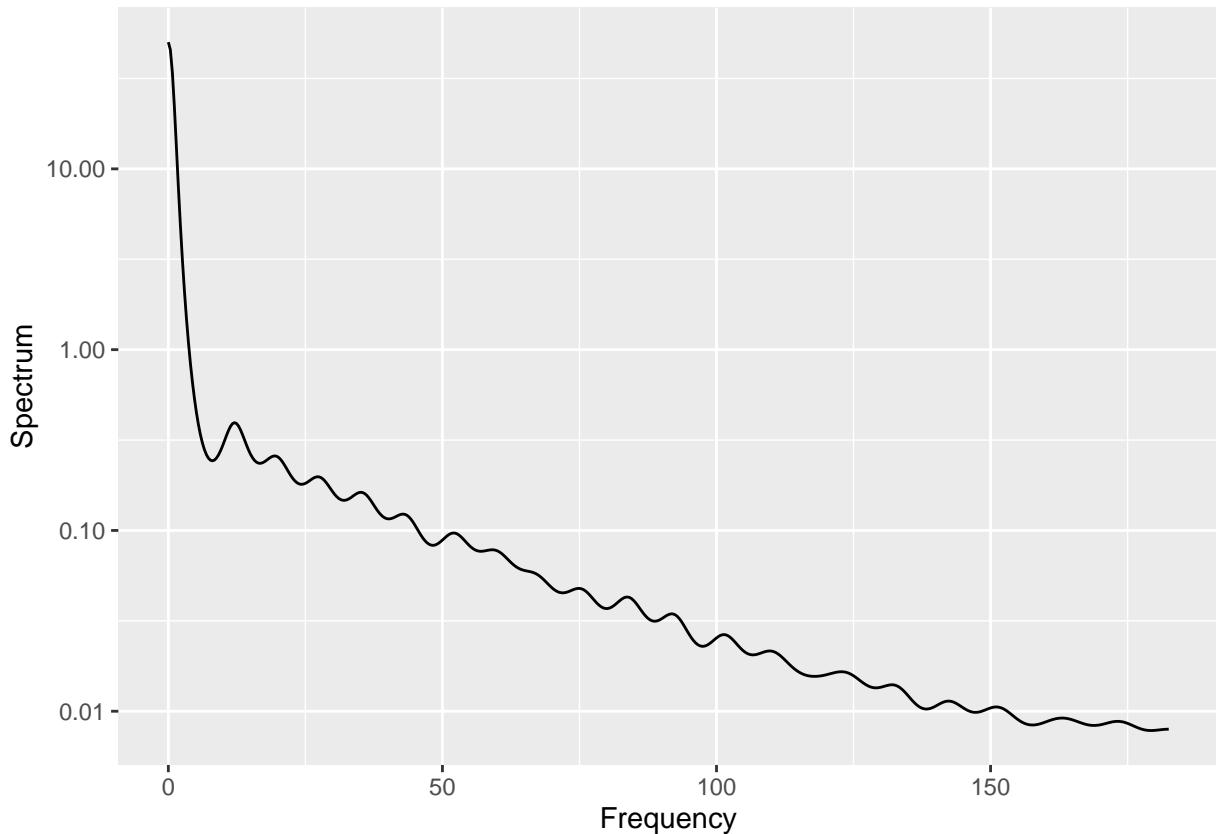


```
autoplot(acf(temp, plot = FALSE), conf.int.fill = '#0000FF', conf.int.value  
↪ = 0.8, conf.int.type = 'ma')
```

Series: temp



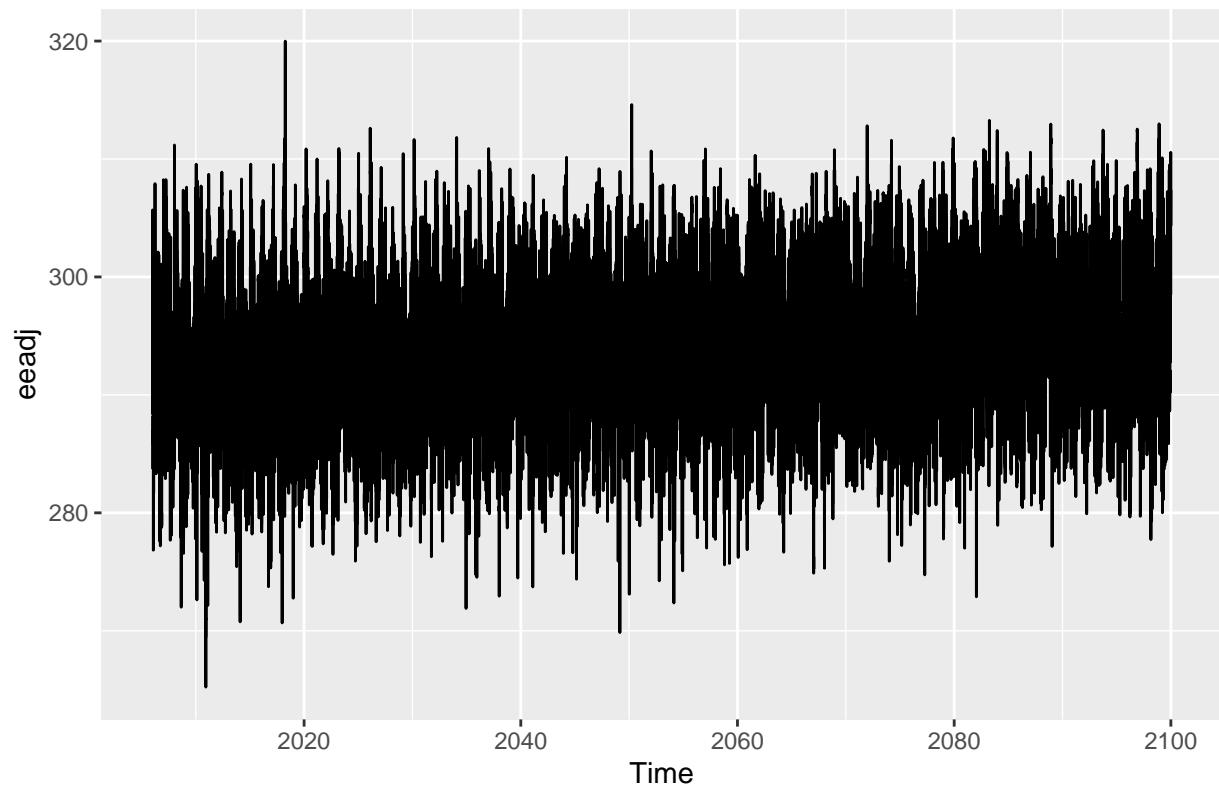
```
autoplot(spec.ar(temp, plot = FALSE))
```



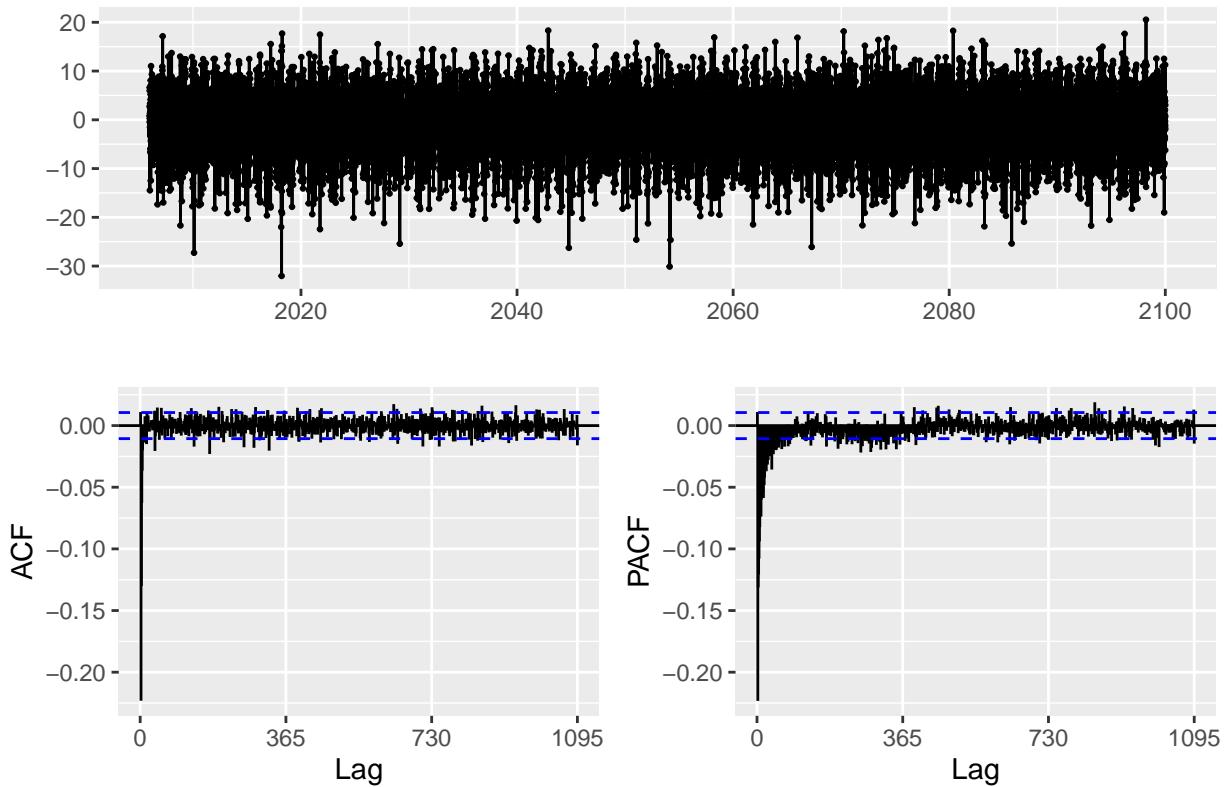
```
library(forecast)
ggttsdiag(auto.arima(temp))
```

```
ggout <- ggfreqplot(temp, freq=100) +
  theme_tufte()
ggsave(ggout, filename = "freqplot.png", width = 11, height = 8.5,
  unit="in", dpi=600)
```

```
library(forecast)
temp %>% stl(s.window='periodic') %>% seasadj() -> eeadj
autoplot(eeadj)
```



```
eadj %>% diff() %>% ggtsdisplay(main="")
```



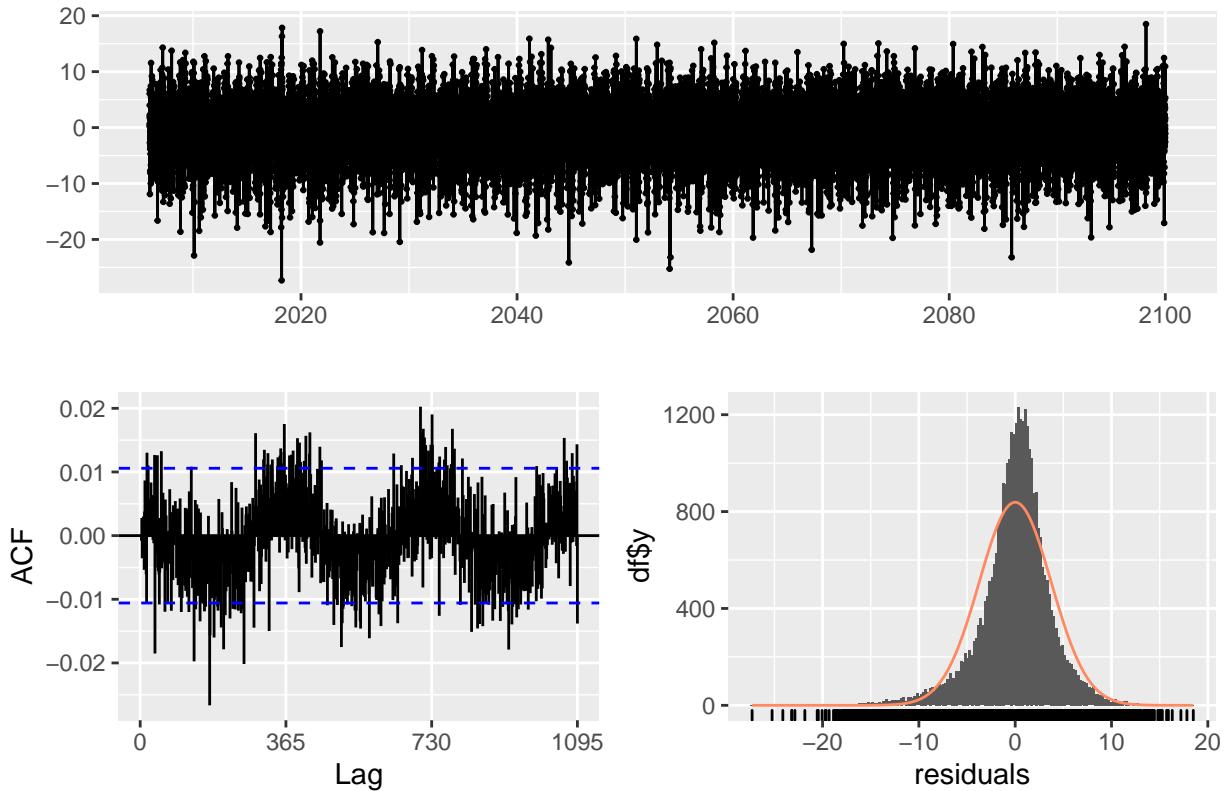
```
(fit <- Arima(eeadj, order=c(5,1,2)))
```

```
## Series: eeadj
## ARIMA(5,1,2)
##
## Coefficients:
##             ar1      ar2      ar3      ar4      ar5      ma1      ma2
##           1.3381  -0.6457  0.1785 -0.0374  0.0114 -1.4619  0.4723
##   s.e.    0.1189   0.1019  0.0293   0.0119  0.0057  0.1188  0.1161
##
## sigma^2 = 13.99:  log likelihood = -93934.25
## AIC=187884.5   AICc=187884.5   BIC=187952

#(fit <- Arima(eeadj, order=c(3,0,1), seasonal=c(0,1,2),
#  lambda=0))

checkresiduals(fit)
```

Residuals from ARIMA(5,1,2)

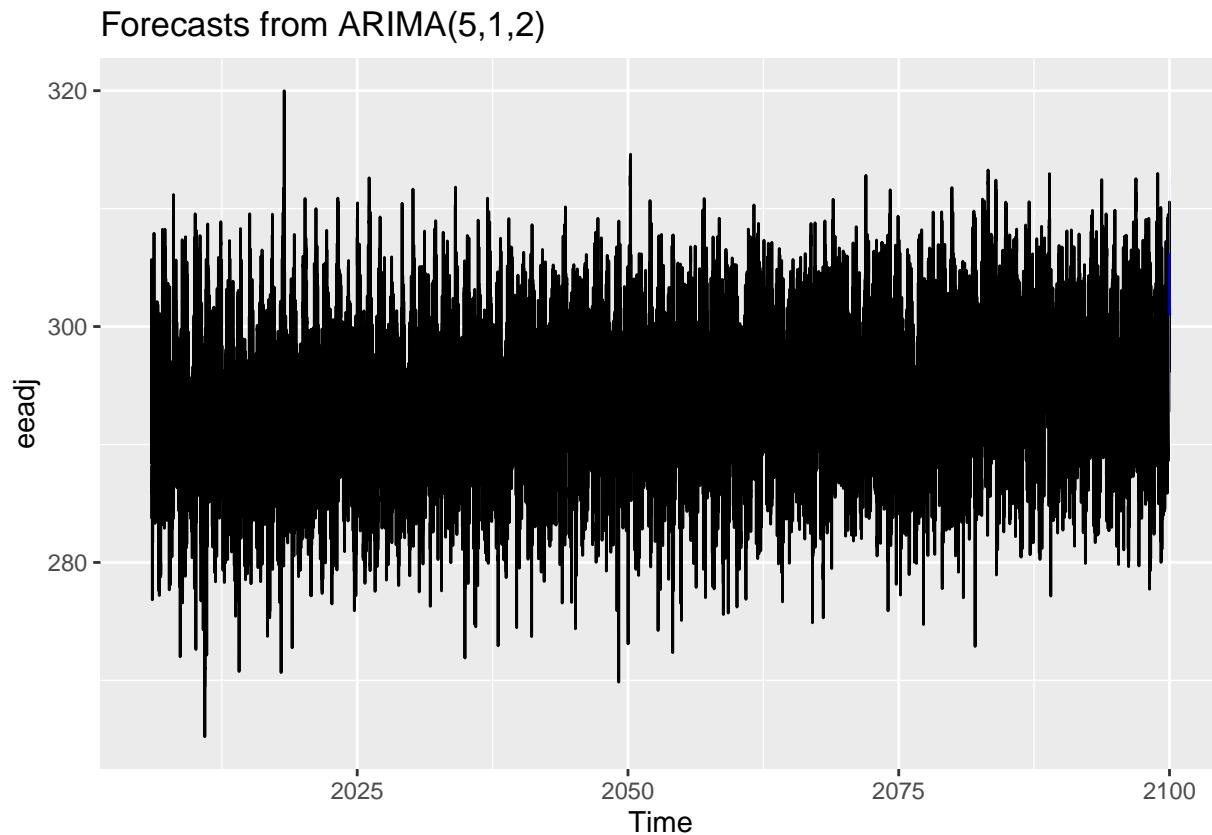


```
##  
## Ljung-Box test  
##  
## data: Residuals from ARIMA(5,1,2)  
## Q* = 1218.6, df = 723, p-value < 2.2e-16  
##  
## Model df: 7. Total lags used: 730
```

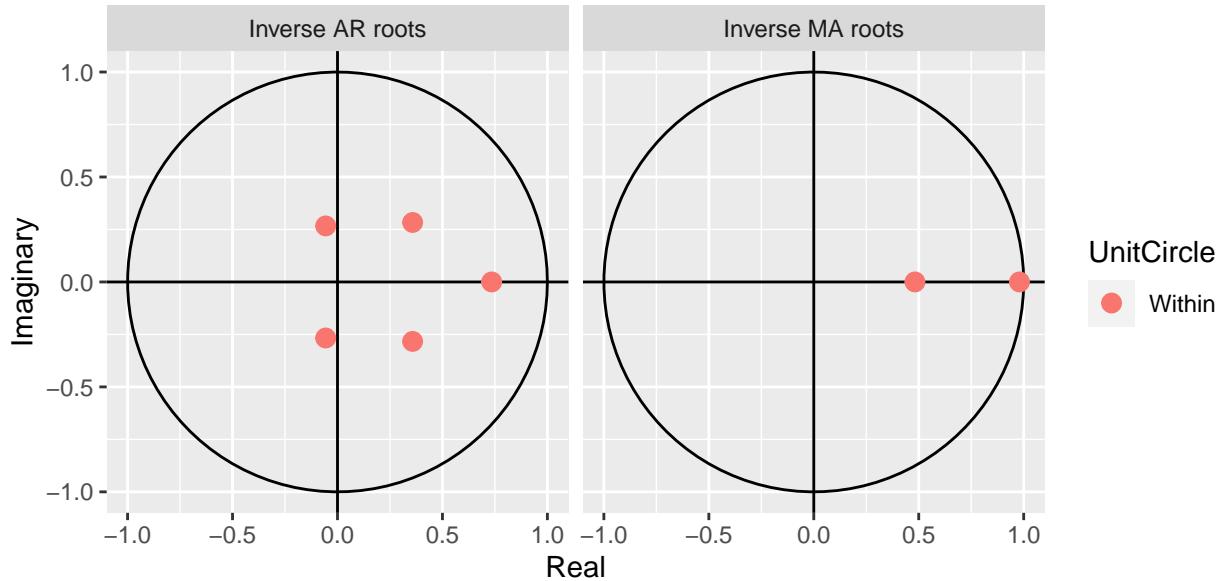
```
auto.arima(eeadj)
```

```
## Series: eeadj  
## ARIMA(5,1,2)  
##  
## Coefficients:  
##          ar1      ar2      ar3      ar4      ar5      ma1      ma2  
##          1.3381  -0.6457  0.1785  -0.0374  0.0114  -1.4619  0.4723  
## s.e.   0.1189   0.1019  0.0293   0.0119  0.0057   0.1188  0.1161
```

```
##  
## sigma^2 = 13.99: log likelihood = -93934.25  
## AIC=187884.5    AICc=187884.5    BIC=187952  
  
fit %>% forecast(h=12) %>% autoplot()
```



```
autoplot(fit)
```



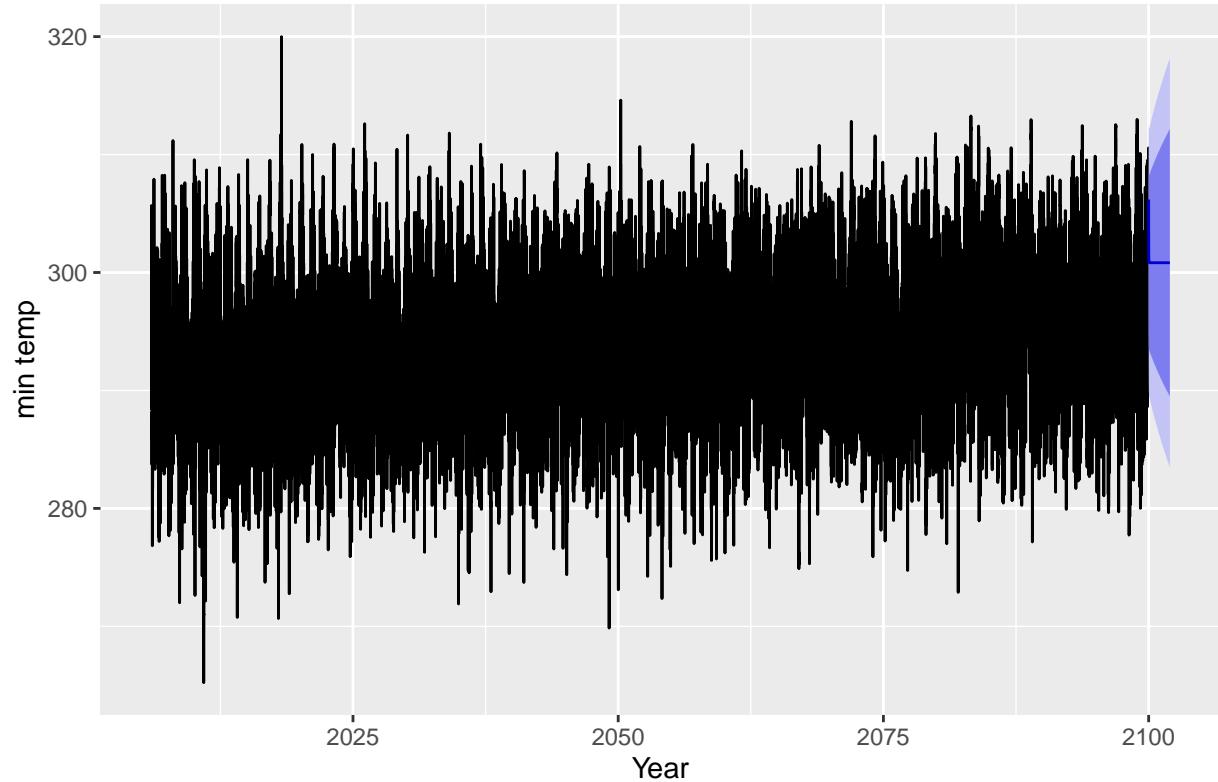
```
length(eeadj)
```

```
## [1] 34310
```

```
interval <- unclass(fit$tsp$interval)
interval <- Filter(function(x) x!=0, interval)

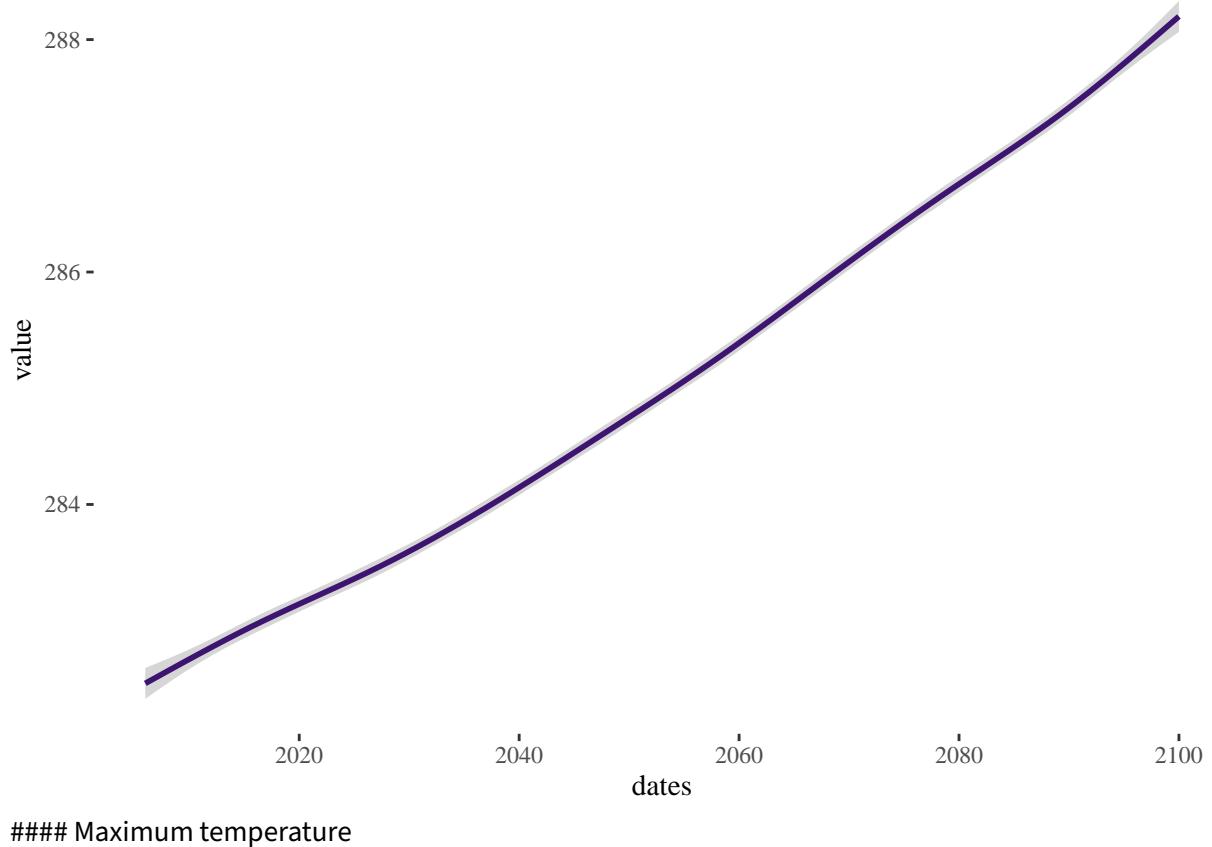
eeadj %>%
  Arima( order=c(5,1,2)) %>%
  forecast() %>%
  autoplot() +
  ylab("min temp") + xlab("Year")
```

Forecasts from ARIMA(5,1,2)



```
ggplot(data= df_min_temp, aes(x = dates, y = value)) +  
  geom_smooth(color=our_purple) +  
  theme_tufte()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

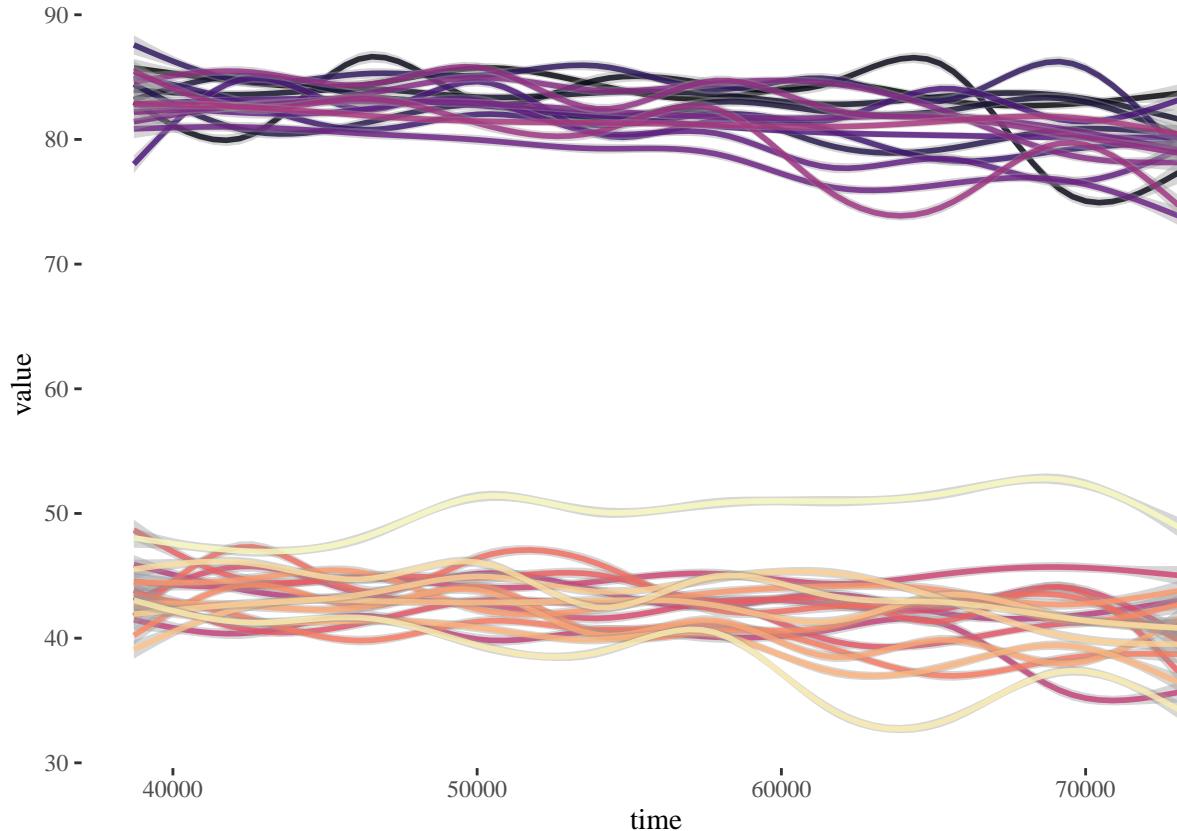


1.6.3 Relative Humidity

```
df_RH <- haskell %>%
  st_drop_geometry() %>%
  select(time, which(startsWith(colnames(haskell), "rhs")))
  ) %>%
  gather(key = "variable", value = "value", -time)

all_climate_projections_RH <- ggplot(data= df_RH, aes(x = time, y = value,
  color = variable)) +
  #geom_line()+
  scale_colour_viridis_d(option="A", alpha = 0.8) +
  geom_smooth() +
  theme_tufte() +
  theme(legend.position = "none")
```

all_climate_projections_RH



```
df_RH_min <- haskell %>%
  st_drop_geometry() %>%
  select(time, which(startsWith(colnames(haskell), "rhsmin"))) %>%
  gather(key = "variable", value = "value", -time)
head(df_RH_min)
```

1.6.3.1 Minimum RH

```
##      time                  variable value
## 1 38716 rhsmin_BNU-ESM_r1i1p1_rcp85   100
## 2 38717 rhsmin_BNU-ESM_r1i1p1_rcp85    79
## 3 38718 rhsmin_BNU-ESM_r1i1p1_rcp85    54
## 4 38719 rhsmin_BNU-ESM_r1i1p1_rcp85    67
```

```
## 5 38720 rhsmin_BNU-ESM_r1i1p1_rcp85      70
## 6 38721 rhsmin_BNU-ESM_r1i1p1_rcp85      63
```

```
df_RH_max <- haskell %>%
  st_drop_geometry() %>%
  select(time, which(startsWith(colnames(haskell), "rhsmax"))) %>%
  gather(key = "variable", value = "value", -time)
head(df_RH_max)
```

1.6.3.2 Maximum RH

```
##   time           variable value
## 1 38716 rhsmax_BNU-ESM_r1i1p1_rcp85 100
## 2 38717 rhsmax_BNU-ESM_r1i1p1_rcp85 100
## 3 38718 rhsmax_BNU-ESM_r1i1p1_rcp85 100
## 4 38719 rhsmax_BNU-ESM_r1i1p1_rcp85  95
## 5 38720 rhsmax_BNU-ESM_r1i1p1_rcp85 100
## 6 38721 rhsmax_BNU-ESM_r1i1p1_rcp85 100
```

1.6.4 Precipitation

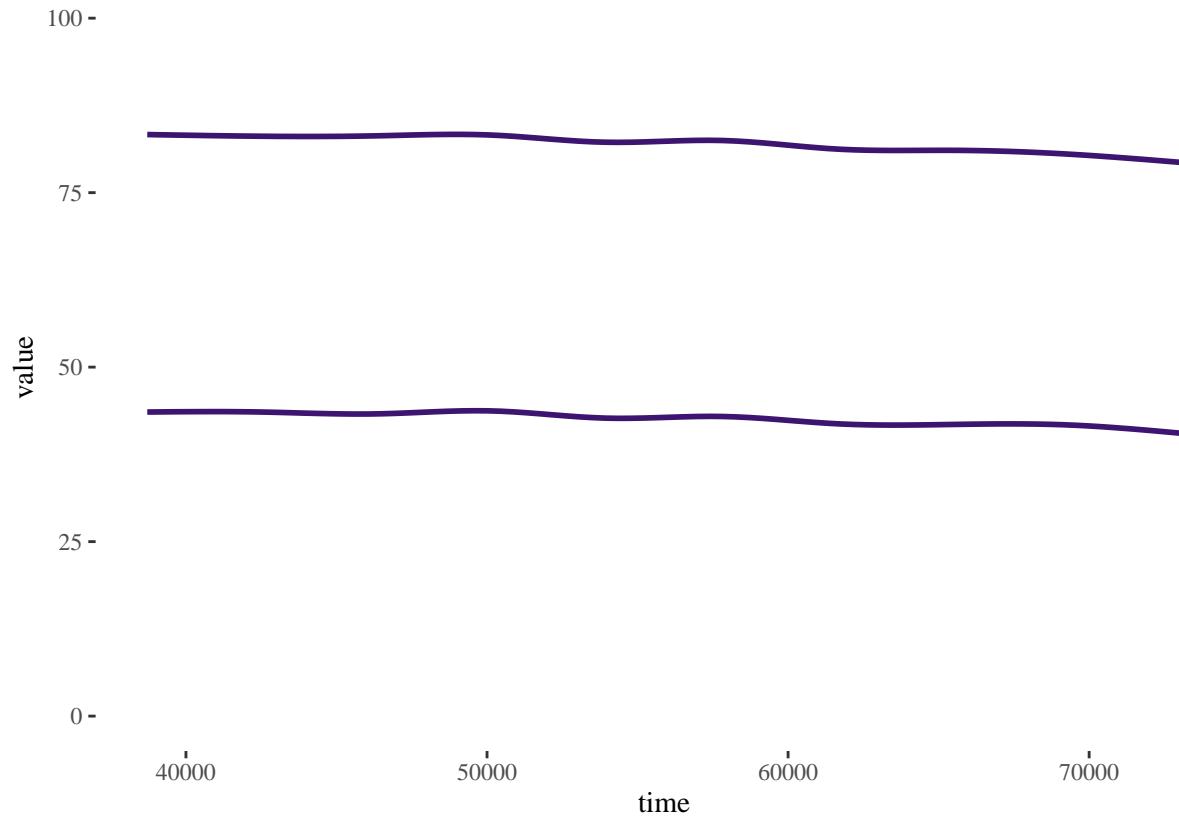
```
df_precip <- haskell_posix %>%
  st_drop_geometry() %>%
  select(dates, which(startsWith(colnames(haskell), "pr"))) %>%
  gather(key = "variable", value = "value", -dates)
head(df_precip)
```

1.6.4.1 Zero-inflated data

```
##   dates           variable value
## 1 2006-01-01 pr_BNU-ESM_r1i1p1_rcp85  0
## 2 2006-01-02 pr_BNU-ESM_r1i1p1_rcp85  0
## 3 2006-01-03 pr_BNU-ESM_r1i1p1_rcp85  0
## 4 2006-01-04 pr_BNU-ESM_r1i1p1_rcp85  0
## 5 2006-01-05 pr_BNU-ESM_r1i1p1_rcp85  0
## 6 2006-01-06 pr_BNU-ESM_r1i1p1_rcp85  0
```

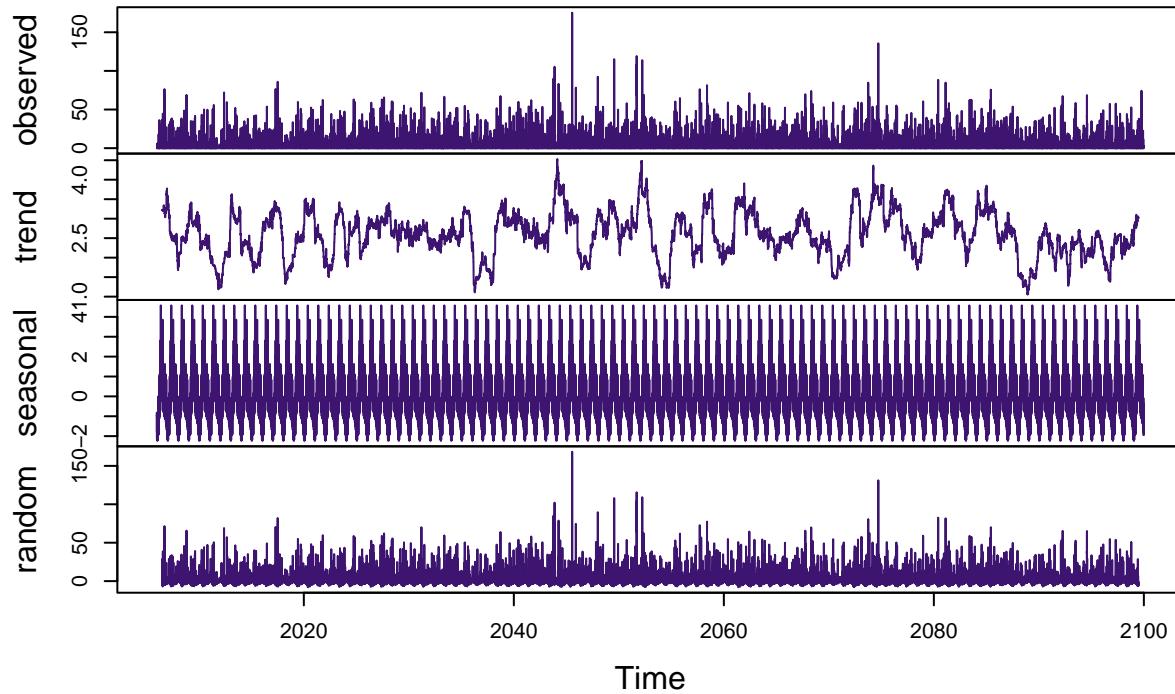
```
all_climate_projections_RH <- ggplot(data= df_RH_min, aes(x = time, y =
  value) )+
  geom_smooth(data= df_RH_min, color=our_purple) +
  geom_smooth(data= df_RH_max, color=our_purple) +
  theme_tufte() +
  ylim(0,100)
```

```
all_climate_projections_RH
```



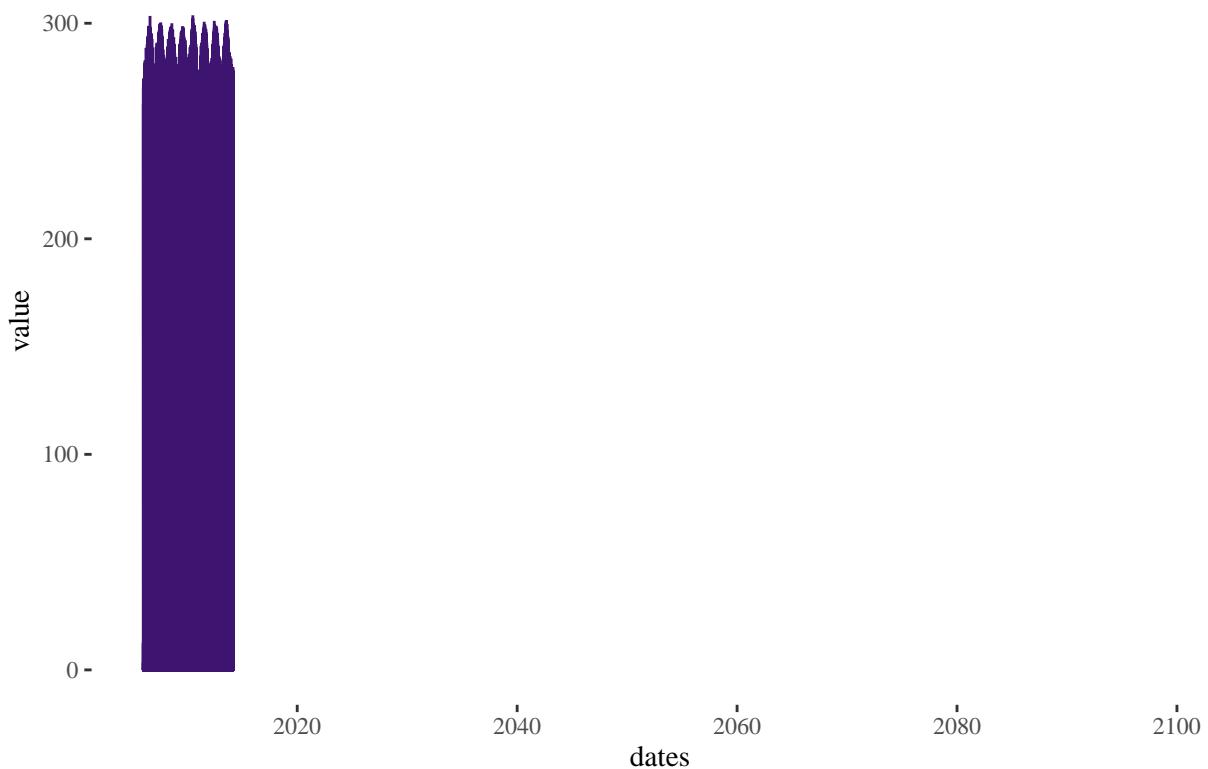
```
precip_ts <- ts(df_precip[,3], frequency = 365, start = c(2006, 1), end =
  c(2099,365))
decomp_precip <- decompose(precip_ts)
plot(decomp_precip, col=our_purple)
```

Decomposition of additive time series

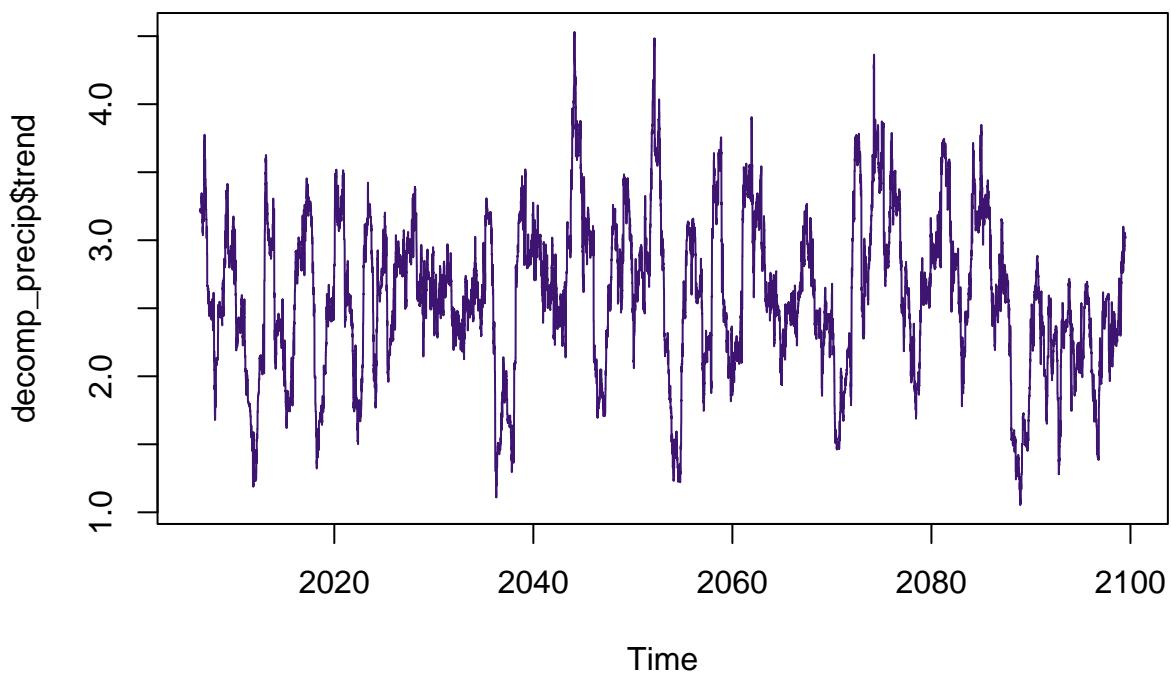


```
all_climate_projections <- ggplot(data= df_precip, aes(x = dates, y =
  ↴ value, color = variable)) +
  geom_line(color=our_purple) +
  theme_tufte()

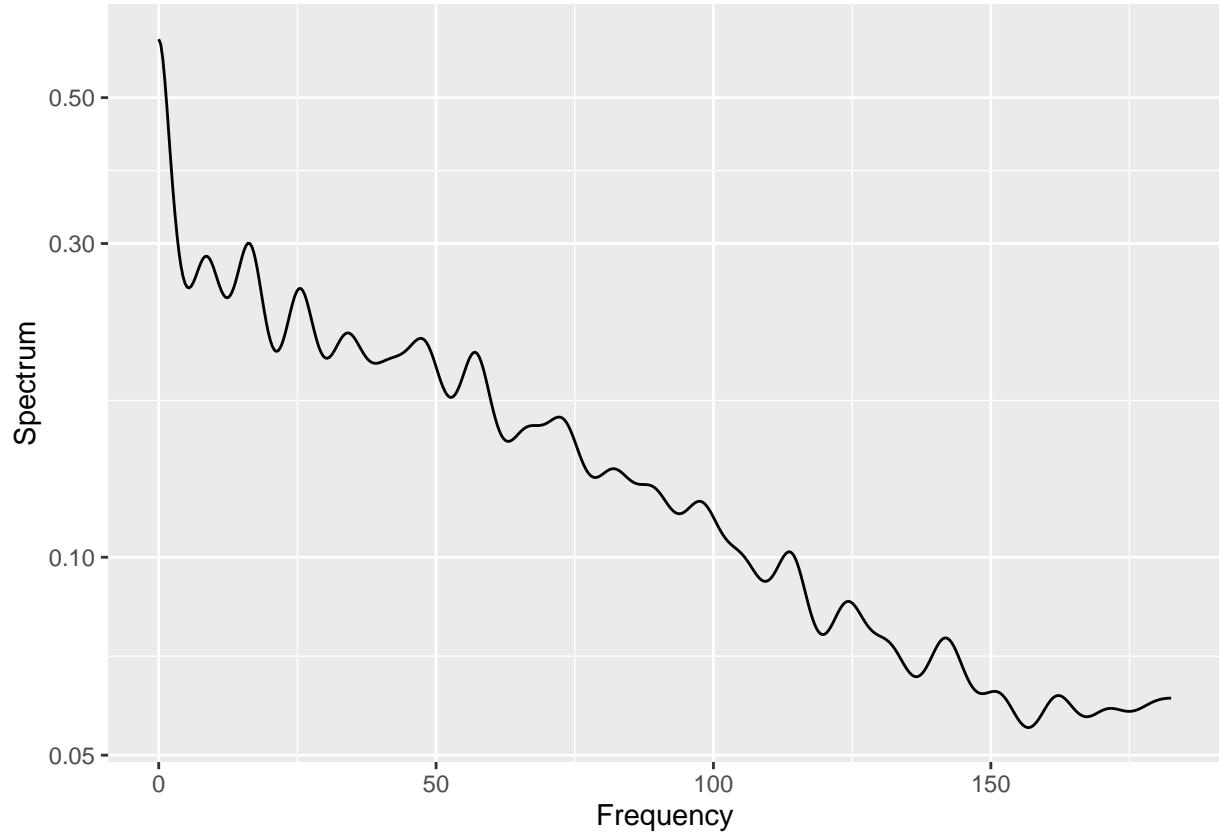
all_climate_projections
```



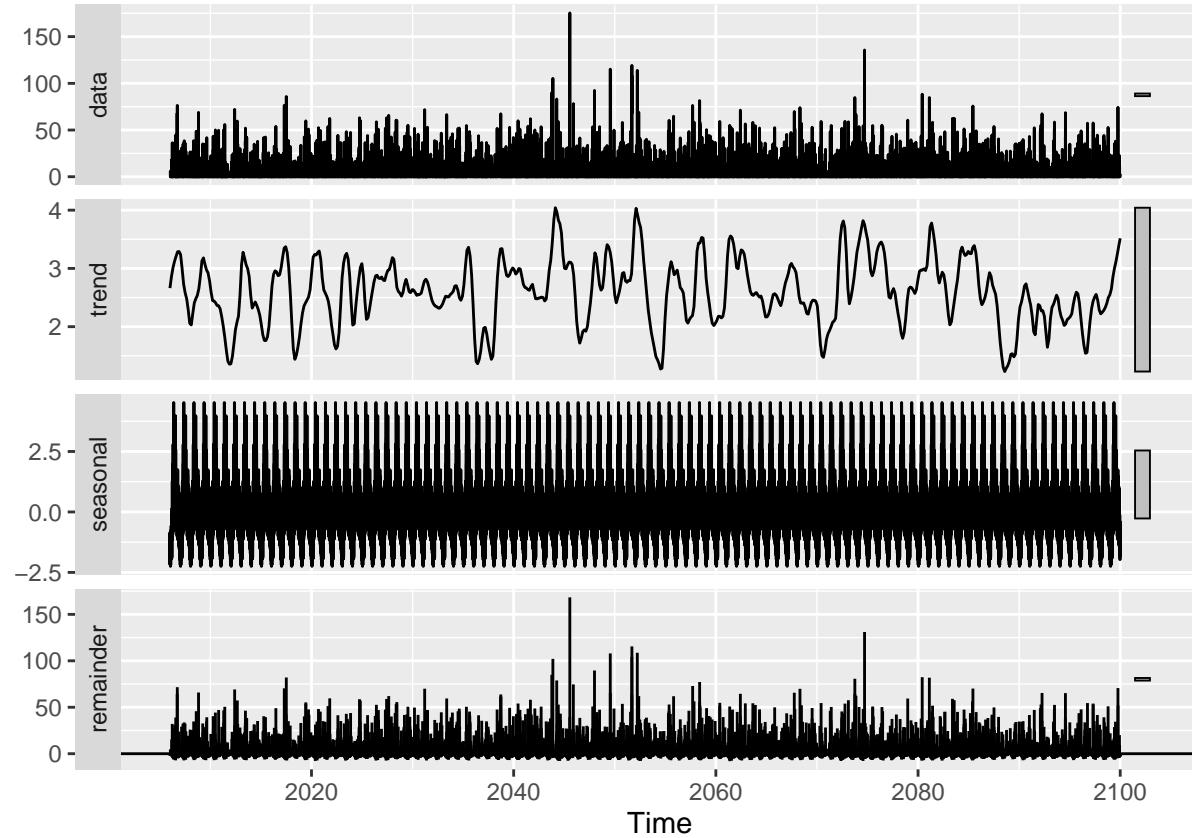
```
plot.ts(decomp_precip$trend, col=our_purple)
```



```
autoplot(spec.ar(precip_ts, plot = FALSE))
```

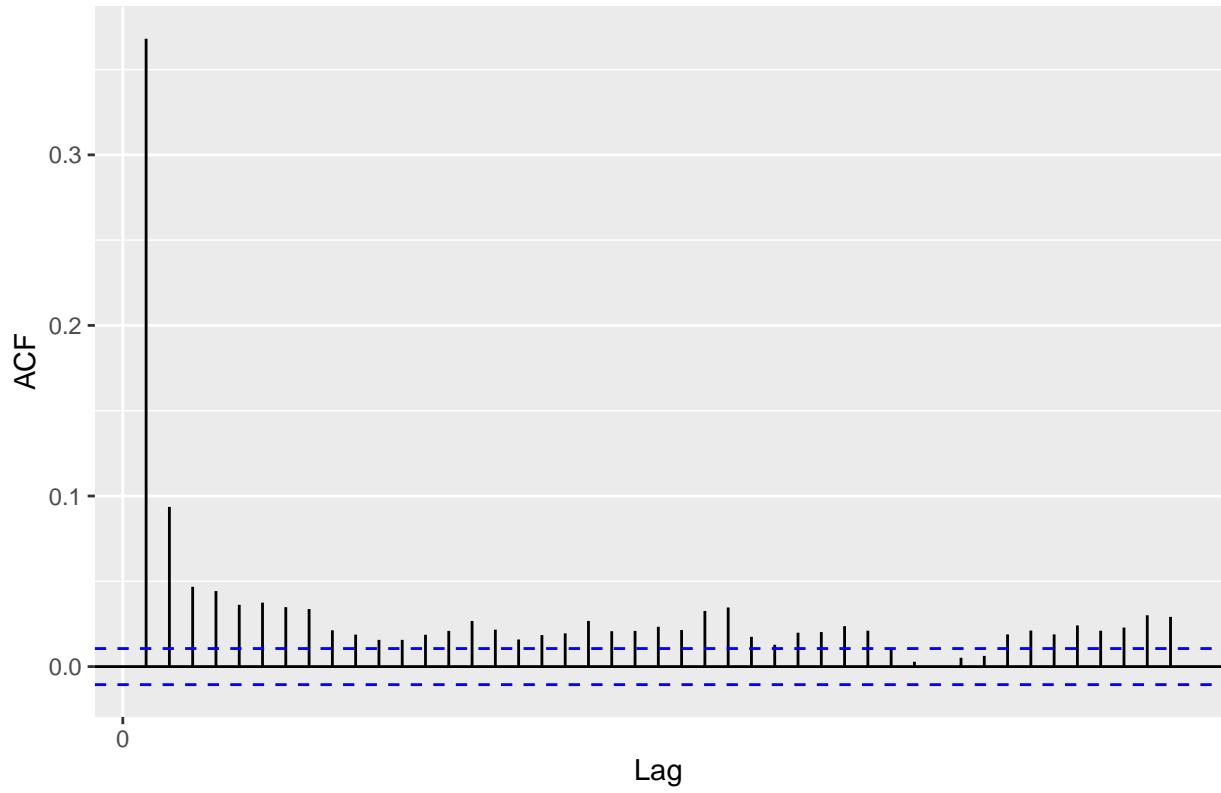


```
autoplot(stl(precip_ts, s.window = 'periodic'), ts.colour = our_purple)
```



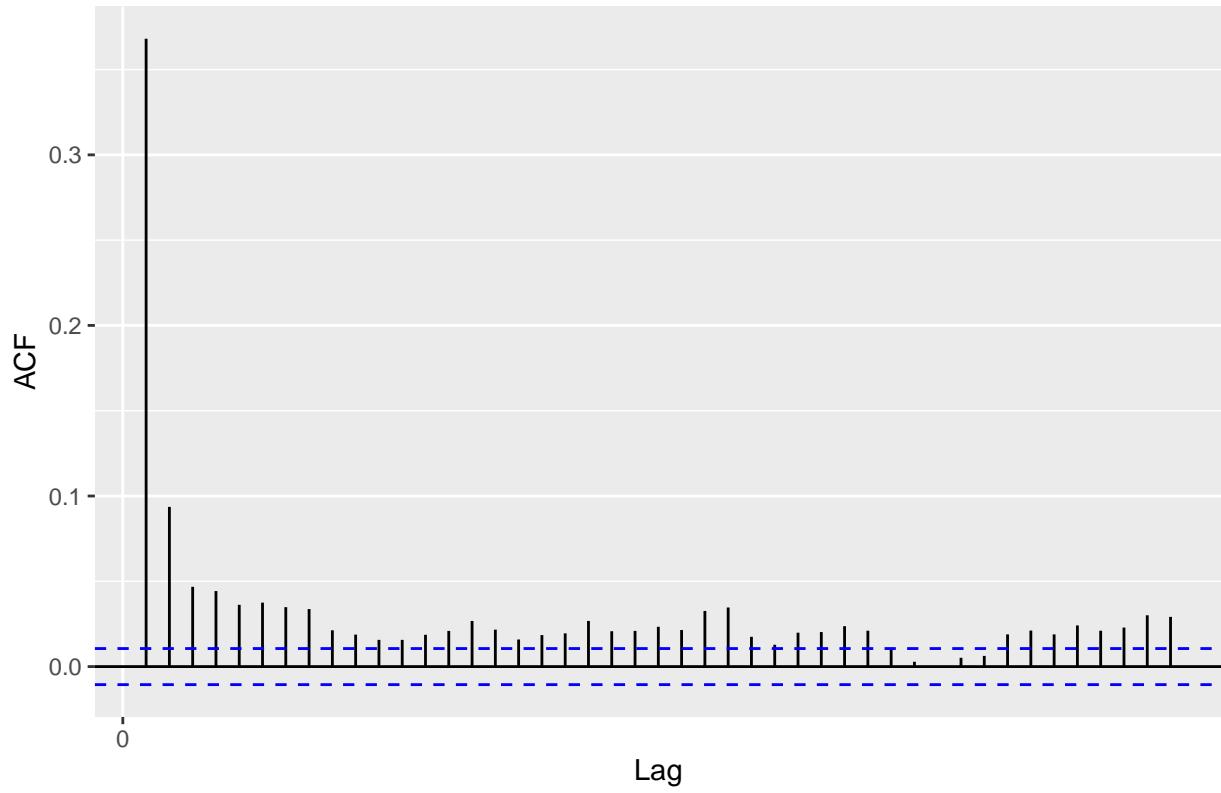
```
autoplot(acf(precip_ts, plot = FALSE))
```

Series: precip_ts

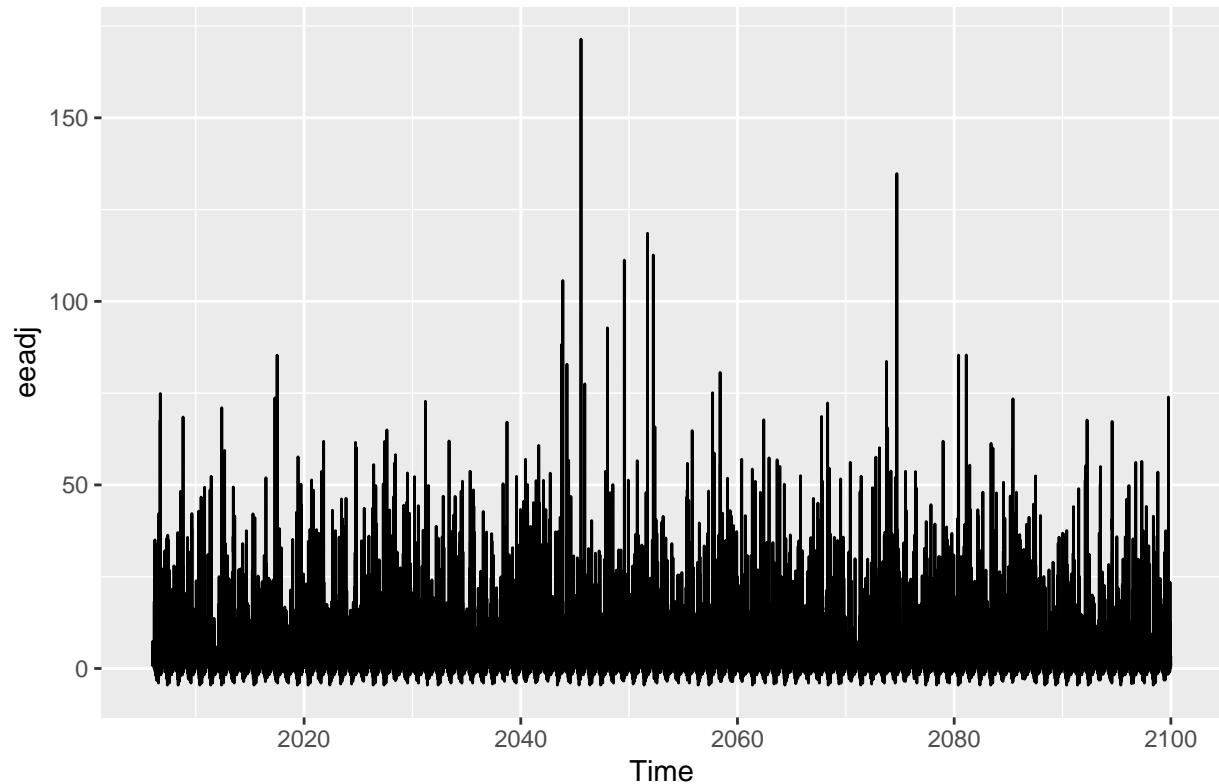


```
autoplot(acf(precip_ts, plot = FALSE), conf.int.fill = '#0000FF',  
        ↵ conf.int.value = 0.8, conf.int.type = 'ma')
```

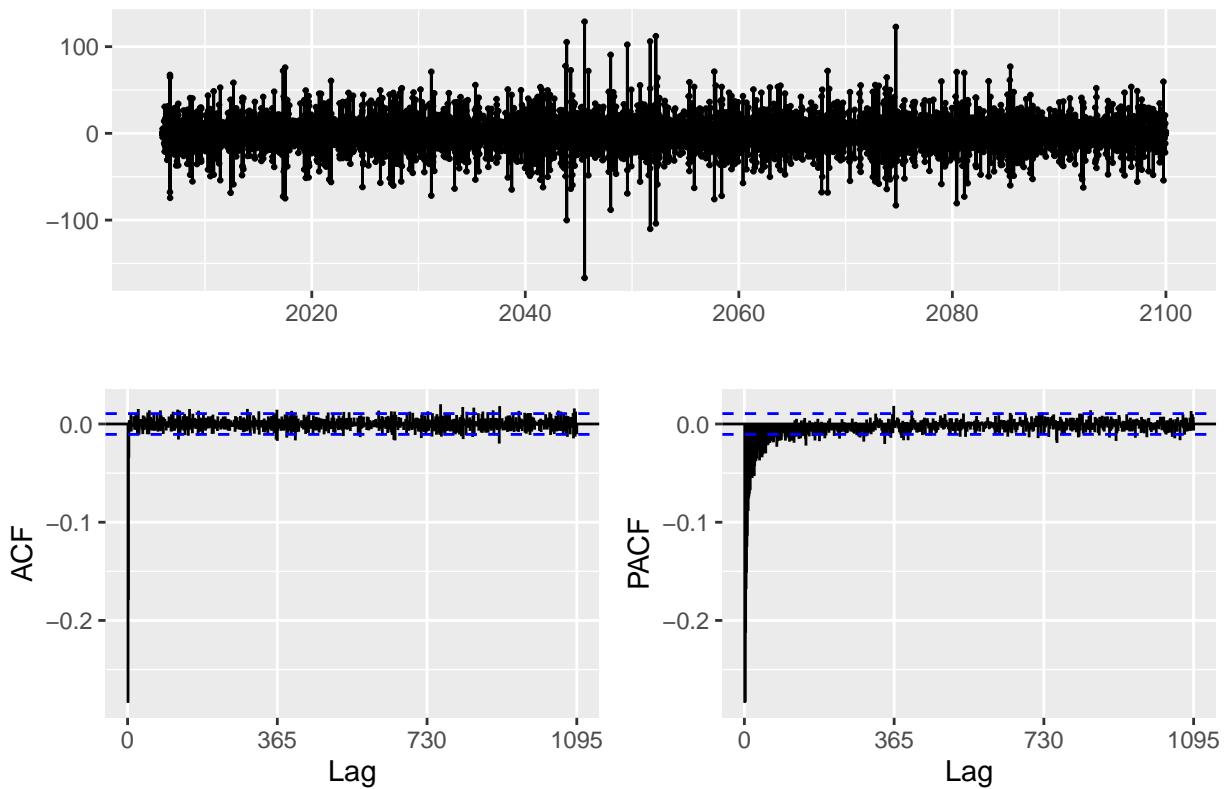
Series: precip_ts



```
precip_ts %>% stl(s.window='periodic') %>% seasadj() -> eeadj  
autoplot(eeadj)
```



```
eadj %>% diff() %>% ggtsdisplay(main="")
```



```
auto.arima(eeadj)
```

```
## Series: eeadj
## ARIMA(3,0,1) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ar3      ma1      mean
##        0.9563  -0.2837  0.0557  -0.5816  2.5955
##  s.e.  0.1281   0.0476  0.0075   0.1282  0.0549
##
## sigma^2 = 43.63:  log likelihood = -113453.1
## AIC=226918.1    AICc=226918.1    BIC=226968.8
```

```
(fit <- Arima(eeadj, order=c(3,0,1)))
```

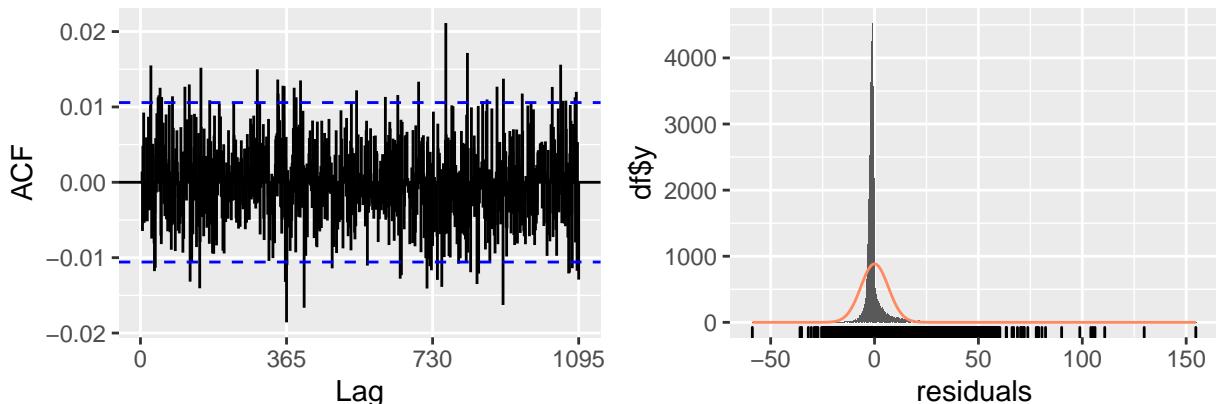
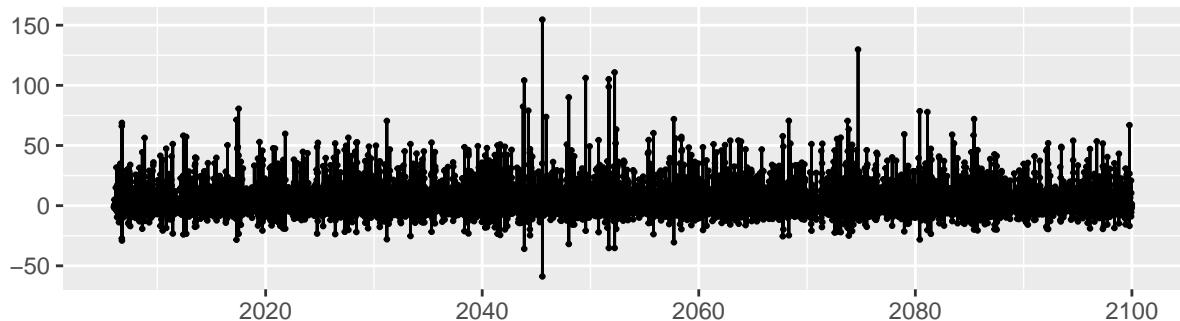
```
## Series: eeadj
## ARIMA(3,0,1) with non-zero mean
##
```

```
## Coefficients:
##          ar1      ar2      ar3      ma1      mean
##        0.9563 -0.2837  0.0557 -0.5816  2.5955
##  s.e.   0.1281  0.0476  0.0075  0.1282  0.0549
## 
## sigma^2 = 43.63: log likelihood = -113453.1
## AIC=226918.1    AICc=226918.1    BIC=226968.8
```

```
#(fit <- Arima(eeadj, order=c(3,0,1), seasonal=c(0,1,2),
#  lambda=0))
```

```
checkresiduals(fit)
```

Residuals from ARIMA(3,0,1) with non-zero mean

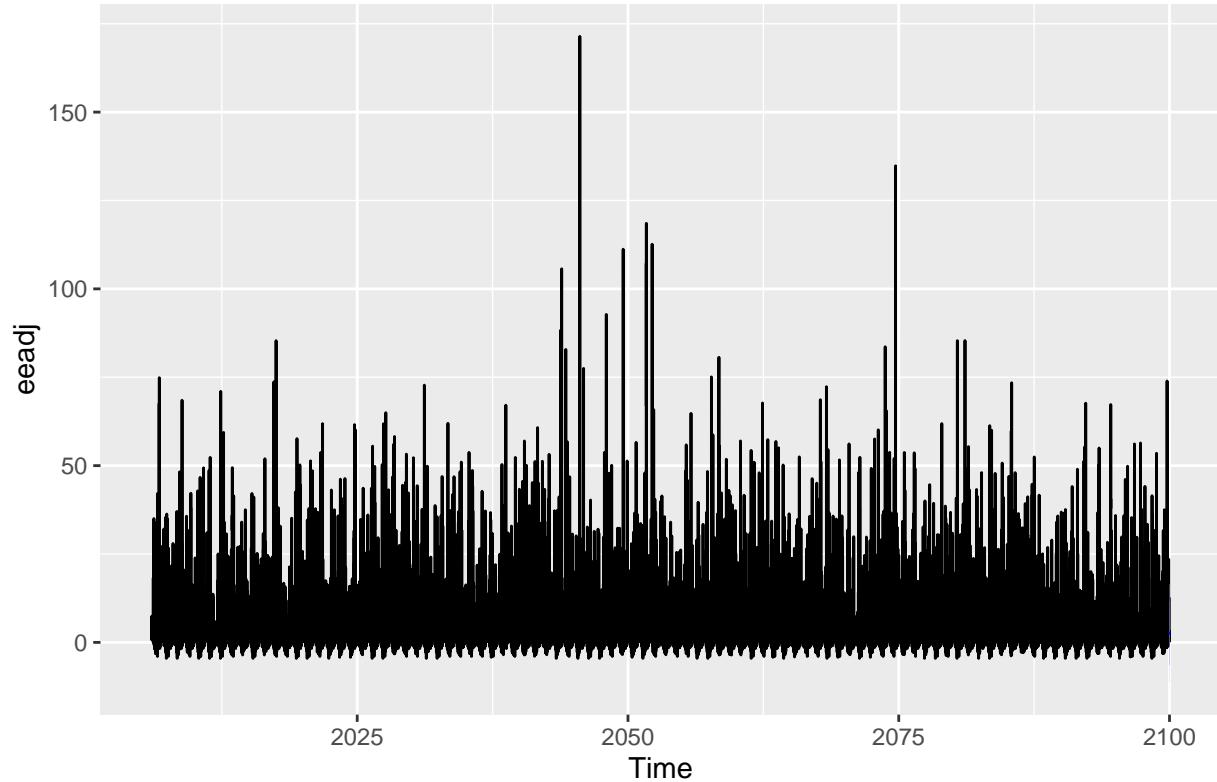


```
## 
## Ljung-Box test
## 
## data: Residuals from ARIMA(3,0,1) with non-zero mean
## Q* = 735.58, df = 726, p-value = 0.3944
```

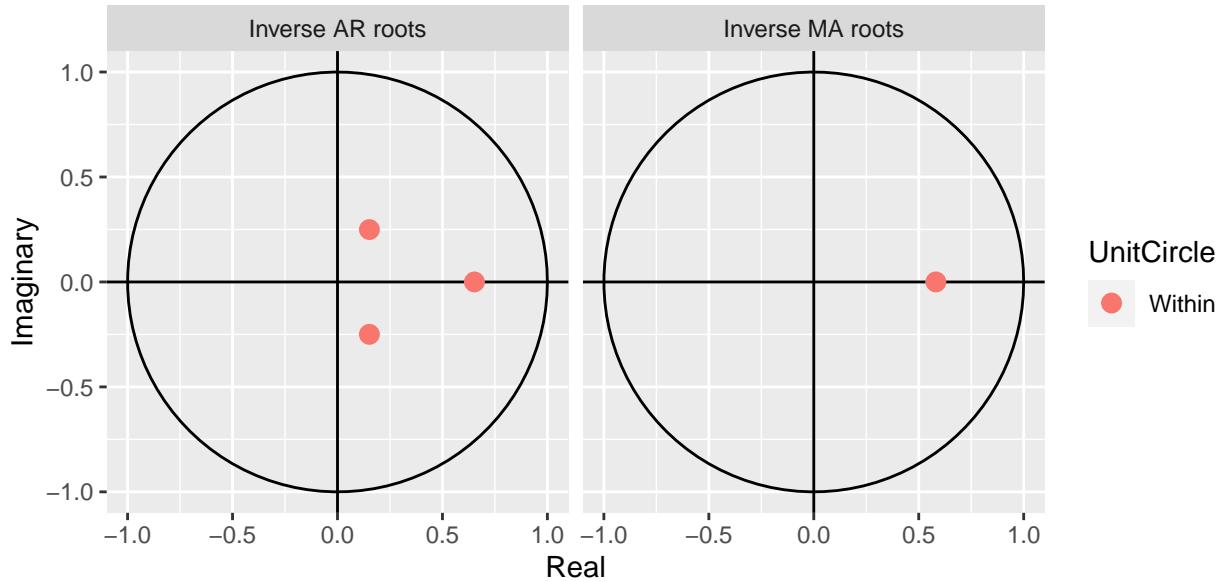
```
##  
## Model df: 4. Total lags used: 730
```

```
fit %>% forecast(h=12) %>% autoplot()
```

Forecasts from ARIMA(3,0,1) with non-zero mean



```
autoplot(fit)
```



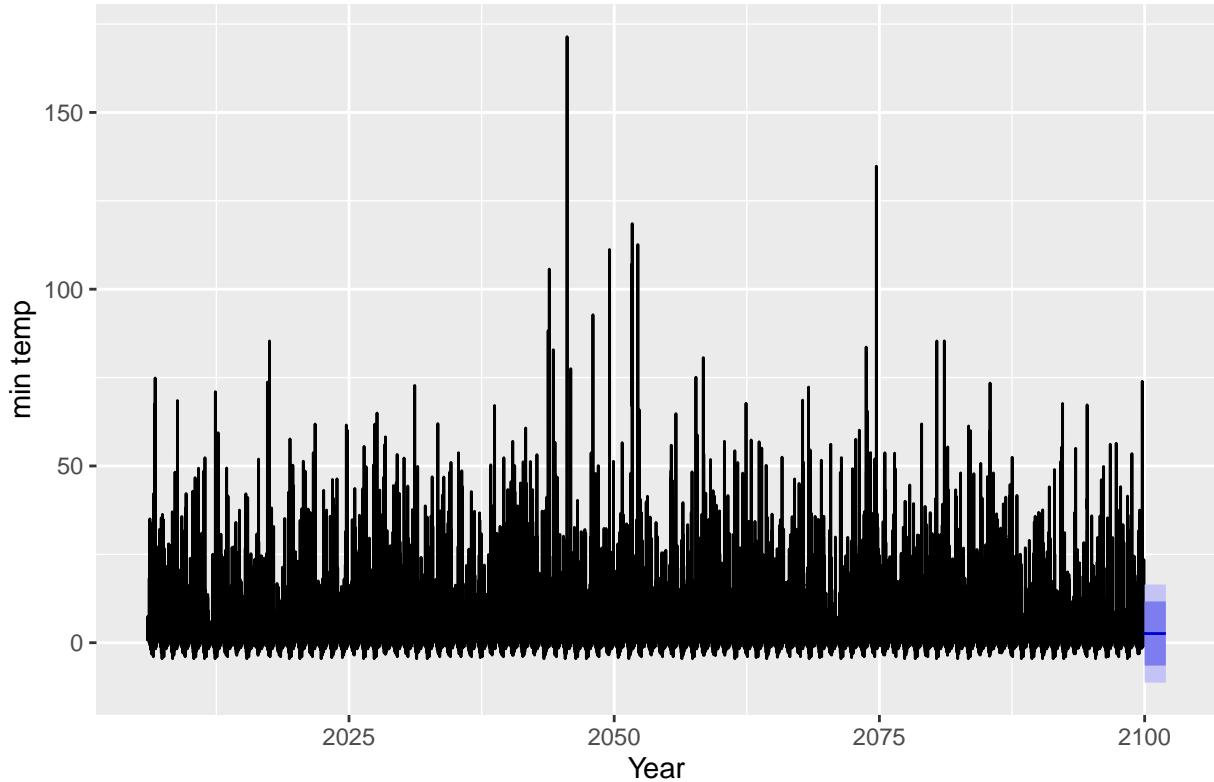
```
length(eeadj)
```

```
## [1] 34310
```

```
interval <- unclass(fit$tsp$interval)
interval <- Filter(function(x) x!=0, interval)

eeadj %>%
  Arima( order=c(3,0,1)) %>%
  forecast() %>%
  autoplot() +
  ylab("min temp") + xlab("Year")
```

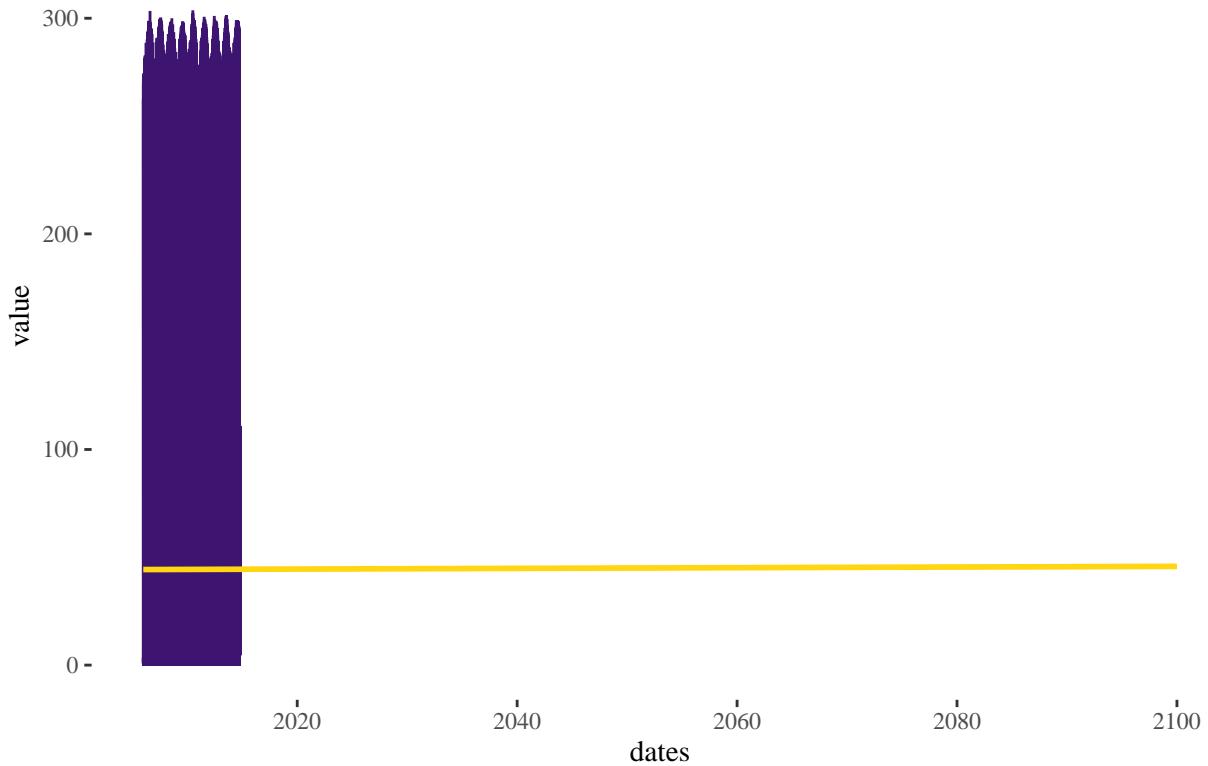
Forecasts from ARIMA(3,0,1) with non-zero mean



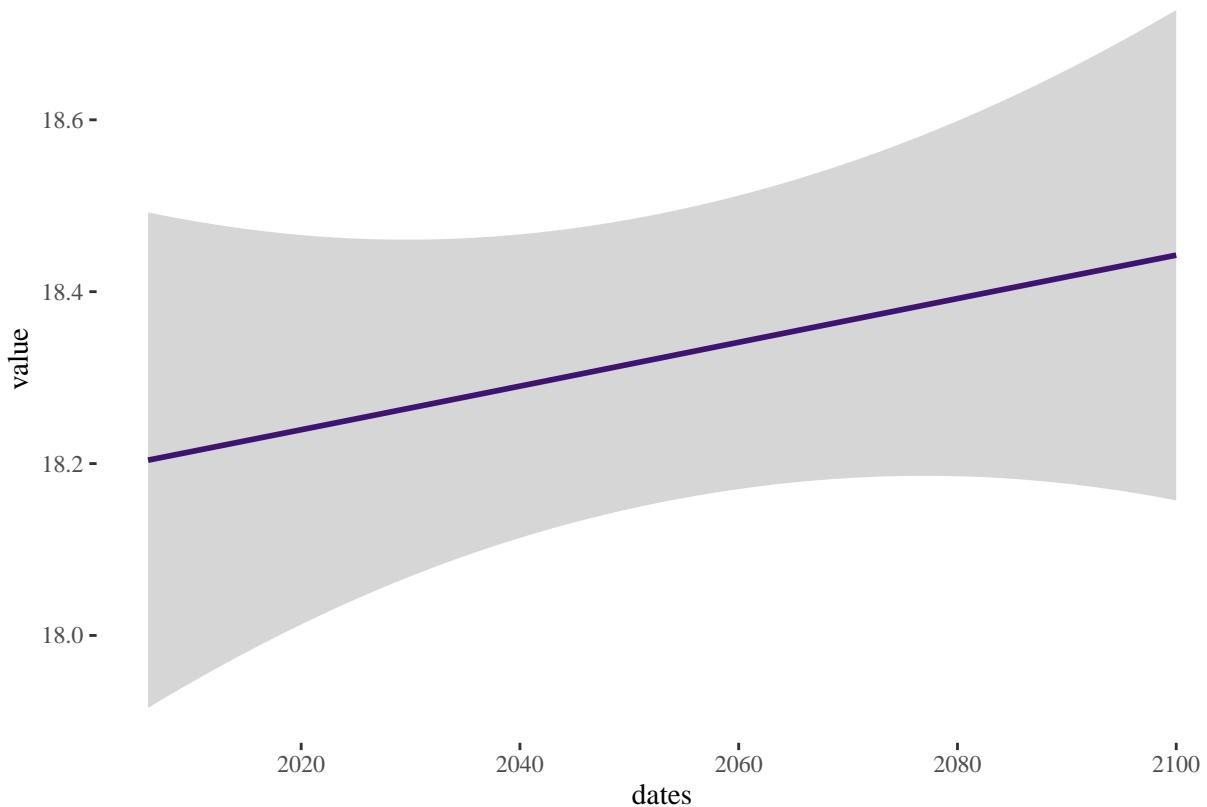
```
df_precip_2 <- df_precip[which(df_precip$value != 0),]

ensemble_climate_projections <- ggplot(data= df_precip_2, aes(x = dates, y
  = value)) +
  geom_line(color=our_purple) +
  geom_smooth(color=our_yellow) +
  theme_tufte()

ensemble_climate_projections
```



```
ggplot(data= df_precip, aes(x = dates, y = value)) +  
  geom_smooth(color=our_purple)+  
  theme_tufte()
```



```
df_precip_quant_95 <- df_precip %>%
  mutate(qu quant = ntile(value, 100)) %>%
  filter(qu quant == 95)

df_precip_quant_96 <- df_precip %>%
  mutate(qu quant = ntile(value, 100)) %>%
  filter(qu quant == 96)

df_precip_quant_97 <- df_precip %>%
  mutate(qu quant = ntile(value, 100)) %>%
  filter(qu quant == 97)

df_precip_quant_98 <- df_precip %>%
  mutate(qu quant = ntile(value, 100)) %>%
  filter(qu quant == 98)

df_precip_quant_99 <- df_precip %>%
  mutate(qu quant = ntile(value, 100)) %>%
  filter(qu quant == 99)
```

```
df_precip_quant_100 <- df_precip %>%
  mutate(quant = ntile(value, 100)) %>%
  filter(quant == 100)

df_precip_quant_1000 <- df_precip %>%
  mutate(quant = ntile(value, 1000)) %>%
  filter(quant == 1000)

df_precip_quant_100000 <- df_precip %>%
  mutate(quant = ntile(value, 100000)) %>%
  filter(quant == 100000)

ggplot(data= df_precip, aes(x = dates, y = value)) +
  geom_smooth(data= df_precip, color=our_purple) +
  geom_smooth(data= df_precip_quant_95, color=our_purple) +
  geom_smooth(data= df_precip_quant_96, color=our_purple) +
  geom_smooth(data= df_precip_quant_97, color=our_purple) +
  geom_smooth(data= df_precip_quant_98, color=our_purple) +
  geom_smooth(data= df_precip_quant_99, color=our_purple) +
  geom_smooth(data= df_precip_quant_100, color=our_purple) +
  geom_smooth(data= df_precip_quant_1000, color=our_purple) +
  geom_smooth(data= df_precip_quant_100000, color=our_purple) +
  theme_tufte()
```

