

Data Application 2

Elizabeth Chase

12/7/2021

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(HPR)
library(mgcv)

## Loading required package: nlme
##
## Attaching package: 'nlme'
##
## The following object is masked from 'package:dplyr':
##
##     collapse
##
## This is mgcv 1.8-34. For overview type 'help("mgcv-package")'.

library(gamlss)

## Loading required package: splines
## Loading required package: gamlss.data
##
## Attaching package: 'gamlss.data'
##
## The following object is masked from 'package:datasets':
##
##     sleep
##
## Loading required package: gamlss.dist
## Loading required package: MASS
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
```

```

## Loading required package: parallel
## ***** GAMLSS Version 5.3-4 *****
## For more on GAMLSS look at https://www.gamlss.com/
## Type gamlssNews() to see new features/changes/bug fixes.
library(kableExtra)

##
## Attaching package: 'kableExtra'
## The following object is masked from 'package:dplyr':
##
##     group_rows
pca_overall <- read_delim("prostate_incidence.txt", delim = "\t", col_names= FALSE)

##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_number(),
##   X4 = col_number()
## )

pca_overall <- pca_overall[-1,]
colnames(pca_overall) <- c("Year", "Rate", "Cases", "Popn")

pca_afa <- read_delim("pca_afa_men.txt", delim = "\t", col_names= FALSE)

##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_number(),
##   X4 = col_number()
## )

pca_afa <- pca_afa[-1,]
colnames(pca_afa) <- c("Year", "Rate", "Cases", "Popn")

pca_nonafa <- data.frame("Year" = pca_overall$Year, "Cases" = pca_overall$Cases-pca_afa$Cases, "Popn" =
pca_afa$Black <- 1

pca_final <- rbind(pca_nonafa, dplyr::select(pca_afa, -Rate))
pca_final$Rate <- 100000*(pca_final$Cases/pca_final$Popn)

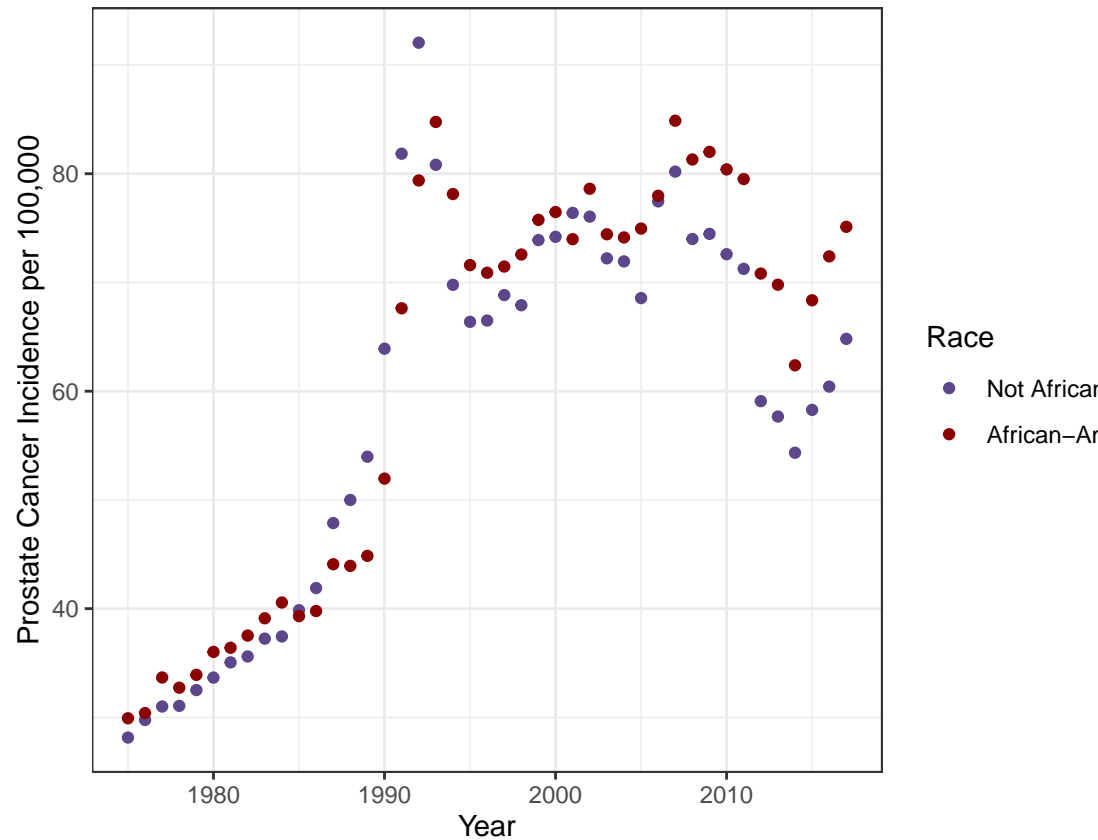
pca_final$Year <- pca_final$Year + 1974

```

Descriptives

This dataset contains 86 observations on the incidence of prostate cancer in the United States on African-American men and non-African American men, 1975-2017. In this case, observations are equally spaced,

and there are two observations at each timepoint (one for African-American men; one for men who aren't African-American).



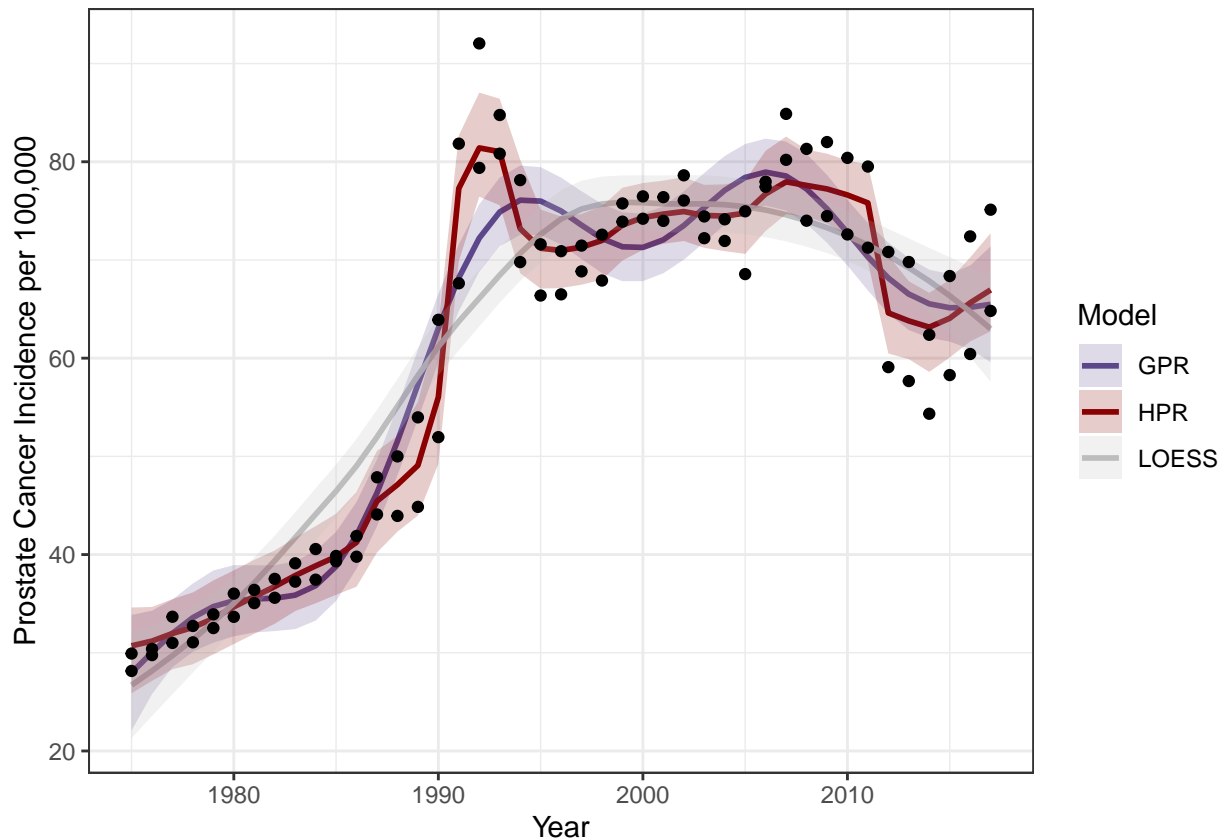
Here's what the data look like:

Model Results

First, a model that does not include race information:

```
## Running MCMC with 4 sequential chains...
## Chain 1 Informational Message: The current Metropolis proposal is about to be rejected because of the
## Chain 1 Exception: normal_lpdf: Location parameter[2] is inf, but must be finite! (in '/var/folders/
## Chain 1 If this warning occurs sporadically, such as for highly constrained variable types like covar
## Chain 1 but if this warning occurs often then your model may be either severely ill-conditioned or m
## Chain 1
## Chain 1 finished in 27.8 seconds.
## Chain 2 finished in 41.4 seconds.
## Chain 3 finished in 26.9 seconds.
## Chain 4 finished in 41.8 seconds.
##
## All 4 chains finished successfully.
## Mean chain execution time: 34.4 seconds.
## Total execution time: 138.1 seconds.
##
## Warning: 105 of 8000 (1.0%) transitions ended with a divergence.
```

```
## This may indicate insufficient exploration of the posterior distribution.
## Possible remedies include:
##   * Increasing adapt_delta closer to 1 (default is 0.8)
##   * Reparameterizing the model (e.g. using a non-centered parameterization)
##   * Using informative or weakly informative prior distributions
## Warning: Dropping 'draws_df' class as required metadata was removed.
```



Now, a model that includes race as a categorical predictor:

```
## Running MCMC with 4 sequential chains...
##
## Chain 1 finished in 32.0 seconds.
## Chain 2 finished in 48.9 seconds.
## Chain 3 finished in 44.1 seconds.
## Chain 4 Informational Message: The current Metropolis proposal is about to be rejected because of the
## Chain 4 Exception: normal_lpdf: Location parameter[2] is inf, but must be finite! (in '/var/folders/...'
## Chain 4 If this warning occurs sporadically, such as for highly constrained variable types like covariates,
## Chain 4 but if this warning occurs often then your model may be either severely ill-conditioned or misspecified.
## Chain 4
## Chain 4 Informational Message: The current Metropolis proposal is about to be rejected because of the
## Chain 4 Exception: normal_lpdf: Location parameter[2] is inf, but must be finite! (in '/var/folders/...'
## Chain 4 If this warning occurs sporadically, such as for highly constrained variable types like covariates,
## Chain 4 but if this warning occurs often then your model may be either severely ill-conditioned or misspecified.
```

```
## Chain 4
## Chain 4 Informational Message: The current Metropolis proposal is about to be rejected because of the
## Chain 4 Exception: normal_lpdf: Location parameter[2] is inf, but must be finite! (in '/var/folders/
## Chain 4 If this warning occurs sporadically, such as for highly constrained variable types like covariance
## Chain 4 but if this warning occurs often then your model may be either severely ill-conditioned or misspecified.
## Chain 4
## Chain 4 finished in 47.4 seconds.
##
## All 4 chains finished successfully.
## Mean chain execution time: 43.1 seconds.
## Total execution time: 172.8 seconds.

##
## Warning: 104 of 8000 (1.0%) transitions ended with a divergence.
## This may indicate insufficient exploration of the posterior distribution.
## Possible remedies include:
##   * Increasing adapt_delta closer to 1 (default is 0.8)
##   * Reparameterizing the model (e.g. using a non-centered parameterization)
##   * Using informative or weakly informative prior distributions

## 1 of 8000 (0.0%) transitions hit the maximum treedepth limit of 12 or 2^12-1 leapfrog steps.
## Trajectories that are prematurely terminated due to this limit will result in slow exploration.
## Increasing the max_treedepth limit can avoid this at the expense of more computation.
## If increasing max_treedepth does not remove warnings, try to reparameterize the model.

## Warning: Dropping 'draws_df' class as required metadata was removed.
```

