

# Condition indices calculated for groundfish species from length-weight residuals

*Chris Rooper, Ebbet Siddon, Jennifer Boldt, Jerry Hoff*

*June 4, 2018*

## PURPOSE

The purpose of this package is to provide a set of functions and a template for R Markdown documents for the computation of groundfish condition in Alaska for the ecosystem considerations chapter of the SAFE. The functions can also be used to calculate L-W residuals for other purposes.

## DATA

The example data sets given here are for the EBS, AI and GOA through the 2017 bottom trawl survey years. However, other surveys that collect similar data (e.g. EcoFOCI surveys, RPA surveys and MACE-AT surveys) can use these functions as well. The r markdown script was prepared by E. Siddon and the functions and package compiled by C. Rooper. A more detailed description of the analyses and ecological relevance can be found in a draft manuscript (Boldt, Rooper and Hoff, in prep, that should be completed in late 2018).

The example data set included in this package is called `lwdata` and is specimen data (length and weight specimens) for all species where this data was collected on RACE bottom trawl surveys. The data was initially extracted from RACEBASE tables (including HAUL, SPECIMEN, SPECIES and STRATA tables). Other sources of data work fine as well, the key points are to have individual lengths and weights, a species identifier, a strata identifier (if you want some sort of by-area calculations) and a year. Here is what the RACEBASE data look like.

```
data("lwdata")
#create a year variable from the cruise data
lwdata["YEAR"]<-round(lwdata["CRUISE"]/100,digits=1)
pander::pandoc.table(lwdata[1:6,])
```

```
##
## -----
##   &nbsp; HAUL   VESSEL   CRUISE   SPECIES_CODE   REGION   START_TIME
## -----
##   **827**    1      88    200101      21720      BS    5/31/2001 6:34:55
##
##   **828**    1      88    200101      21720      BS    5/31/2001 6:34:55
##
##   **829**    1      88    200101      21720      BS    5/31/2001 6:34:55
##
##   **830**    1      88    200101      21720      BS    5/31/2001 6:34:55
##
##   **831**    1      88    200101      21720      BS    5/31/2001 6:34:55
##
##   **832**    1      88    200101      21720      BS    5/31/2001 6:34:55
## -----
##
## Table: Table continues below
##
## -----
##   &nbsp; INPFC_AREA   STRATUM   STATIONID   BOTTOM_DEPTH   START_LATITUDE
## -----
##   **827**           10      F-14           36           56.67
##
##   **828**           10      F-14           36           56.67
##
##   **829**           10      F-14           36           56.67
##
##   **830**           10      F-14           36           56.67
```

```
##
## **831**          10          F-14          36          56.67
##
## **832**          10          F-14          36          56.67
## -----
##
## Table: Table continues below
##
## -----
## &nbsp; START_LONGITUDE SPECIMENID SPECIES_NAME COMMON_NAME
## -----
## **827**      -159.7         14      Gadus macrocephalus Pacific cod
##
## **828**      -159.7         7       Gadus macrocephalus Pacific cod
##
## **829**      -159.7         2       Gadus macrocephalus Pacific cod
##
## **830**      -159.7         5       Gadus macrocephalus Pacific cod
##
## **831**      -159.7         3       Gadus macrocephalus Pacific cod
##
## **832**      -159.7         9       Gadus macrocephalus Pacific cod
## -----
##
## Table: Table continues below
##
## -----
## &nbsp; SEX  LENGTH  WEIGHT  CATCH  YEAR
## -----
## **827**    2    310    252    24.34  2001
##
## **828**    1    170    40     24.34  2001
##
## **829**    1    210    82     24.34  2001
##
## **830**    1    220    98     24.34  2001
##
## **831**    1    200    66     24.34  2001
##
## **832**    1    370    482    24.34  2001
## -----
```

## FUNCTIONS

The key function for this analysis is the calculation of residuals from a log-log relationship between length and weight the `lw.resids` function inputs the length and weight data for individual fish and estimates the residuals.

```
lw.resids<-function(length,weight){
  loglength<-log(length)
  logwt<-log(weight)
  lw.res<-lm(logwt~loglength)
  lw.res<-lw.res$residuals
```

```
return(lw.res)}
```

## EXAMPLE

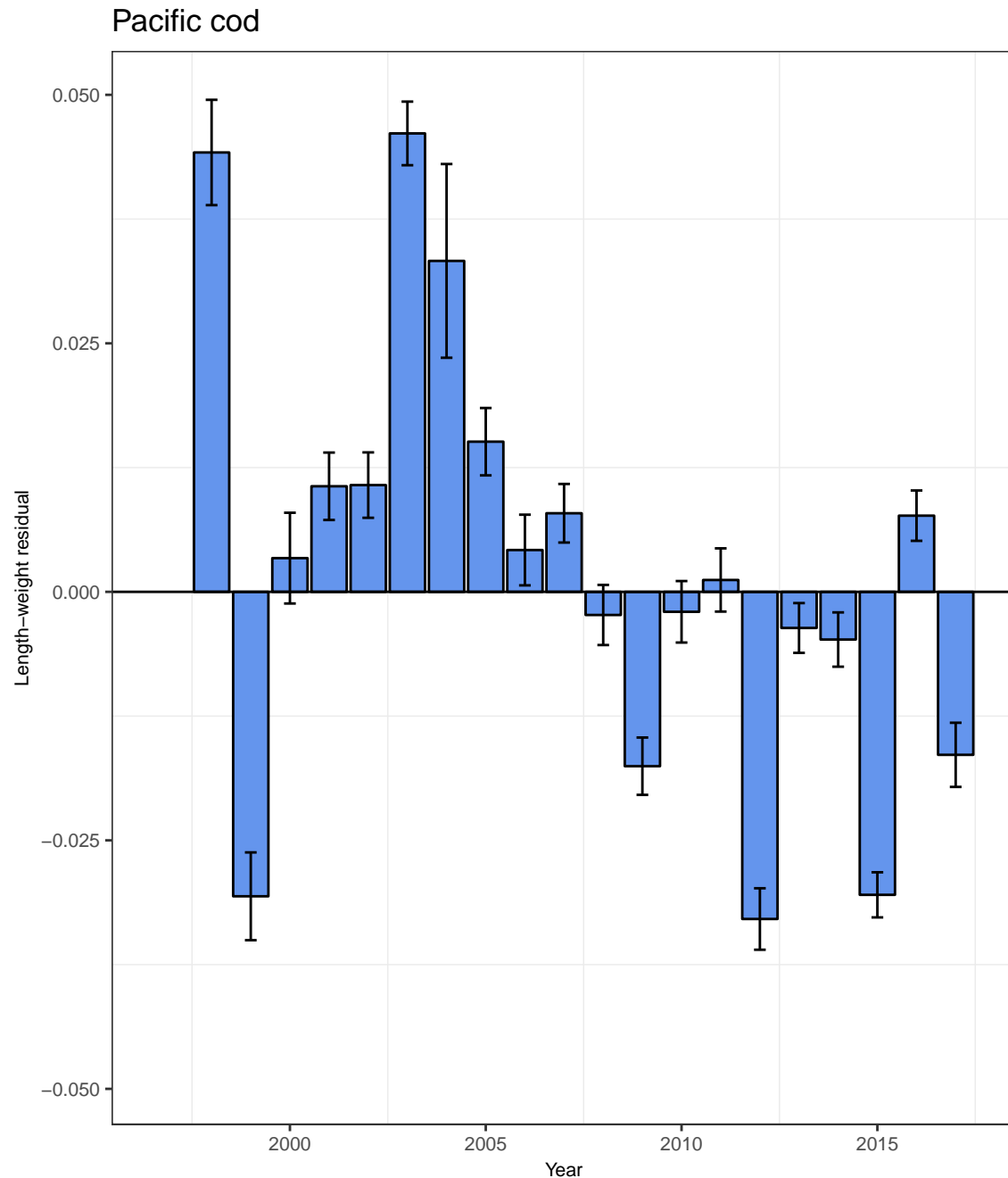
As an example of calculating the condition index, we take Pacific Cod in the eastern Bering Sea. The code is designed to loop through a number of species (e.g. see the EBS\_Groundfish\_CONDITION.Rmd file which computes the index for a group of species used in the Ecosystem Considerations SAFE Chapter). For this example, we are doing only a single species, but have left the architecture for multiple species intact. Also, please note that some of the steps such as the STRATUM definitions are specific to the RACEBASE EBS data and may not be needed for other applications.

The plotting by year is done in ggplot2 using gridExtra. This code can be modified as needed, but basically the mean L-W residuals (and variances) are calculated and put into a bar plot with year as the x-axis and mean residual on the y-axis. We have made the plots into a list object (myplot) so that they can easily be arranged onto a gridded figure.

```
myplot<-list() #Make the empty list
#myplot[]
lwdata_by_year<-array(dim=c(0,6))
colnames(lwdata_by_year)<-c("species","yrs","ymean","yn","ysd","yse")
for(i in 1:length(EBS.species)){ #set up the loop to loop through species
  tempdata<-subset(EBS.lwdata,EBS.lwdata$SPECIES_CODE==EBS.species[i]) #subset the data to the species
  tempdata["residuals"]<-lw.resids(tempdata$LENGTH,tempdata$WEIGHT) # Use the lw.resids function to calculate residuals
  yrs=sort(unique(tempdata$YEAR)) #Sort by year
  ymean=apply(tempdata$residuals,tempdata$YEAR,mean) #Calculate mean by year
  yn=apply(tempdata$residuals,tempdata$YEAR,length) #Count the number of observations by year
  ysd=apply(tempdata$residuals,tempdata$YEAR,sd) #Calculate the sd of the mean by year
  yse=ysd/sqrt(yn) #Calculate the standard error
  data.summary<-data.frame(EBS.species[i],yrs,ymean,yn,ysd,yse) #Put the mean, SE and year into a data frame
  lwdata_by_year<-rbind(lwdata_by_year,data.summary)
  p<-ggplot(data.summary, aes(x = yrs, y = ymean,cex=2) +
    geom_bar(position = position_dodge(), stat="identity", fill="cornflowerblue",col="black") +
    geom_errorbar(aes(ymin=ymean-yse, ymax=ymean+yse),width=0.30) +
    xlim(1996.5, 2017.5)+
    ggtitle(paste(EBS.speciesnames[i])) +
    geom_hline(yintercept=0, color="black")+
    theme_bw() +
    theme(panel.grid.major = element_blank()+
      theme(axis.text.x = element_text(size=8))+
    theme(axis.text.y = element_text(size=8))+
    theme(axis.title.x = element_text(size=8))+
    theme(axis.title.y = element_text(size=8))+
    labs(title = paste(EBS.speciesnames[i]), y = "Length-weight residual", x = "Year")
  #print(p)
  pltName <- paste( EBS.speciesnames[i],"plot", sep = ' ' ) #Name the plot
  myplot[[pltName]]<-p} #Add the plot to the list and loop
```

This line of code does the arranging of the plots on a grid. In this case we have a single species (Pacific cod), so the number of columns (ncol) is set to 1. For the Ecosystem Considerations contributions there are usually 2 columns (ncol=2). It also outputs the figure as a .png to the working directory and outputs the data to a .csv file.

```
## pdf
## 2
```



In this next bit of code, the same procedures are followed for calculating the residuals and producing plots. This time the plots are by strata and year. Plots and data are output at the end.

```
lwdata_by_strata<-array(dim=c(0,7))
colnames(lwdata_by_strata)<-c("species","strata","yrs","ymean","yn","ysd","yse")
#By stratum graphs
for(i in 1:length(EBS.species)){
  tempdata<-subset(EBS.lwdata,EBS.lwdata$SPECIES_CODE==EBS.species[i])
  tempdata["residuals"]<-lw.resids(tempdata$LENGTH,tempdata$WEIGHT)

  ymeans=aggregate(tempdata$residuals,by=list(tempdata$YEAR,tempdata$STRATUM),mean)
```

```

ysd=aggregate(tempdata$residuals,by=list(tempdata$YEAR,tempdata$STRATUM),sd)
yn=aggregate(tempdata$residuals,by=list(tempdata$YEAR,tempdata$STRATUM),length)
yse=ysd$x/sqrt(yn$x)
data.summary<-data.frame(species=EBS.species[i],strata=ymeans$Group.2,yrs=ymeans$Group.1,ymeans=ymeans$
lwdata_by_strata<-rbind(lwdata_by_strata,data.summary)

dat1 <- subset(data.summary,data.summary$ymean>=0)
dat2 <- subset(data.summary,data.summary$ymean< 0)
p2<-ggplot() +
  geom_bar(data = dat1, aes(x=yrs, y=ymean, fill=factor(strata)),stat = "identity",col="black") +
  geom_bar(data = dat2, aes(x=yrs, y=ymean, fill=factor(strata)),stat = "identity",col="black")+
  scale_fill_brewer(palette = "Spectral")+
  geom_hline(yintercept=0, color="black")+
  xlim(1996.5, 2017.5)+
  theme_bw() +
  theme(axis.text.x = element_text(size=8))+
  theme(axis.text.y = element_text(size=8))+
  theme(axis.title.x = element_text(size=8))+
  theme(axis.title.y = element_text(size=8))+
  theme(panel.grid.major = element_blank())+theme(legend.position="right")+
  labs(title = paste(EBS.speciesnames[i]), y = "Length-weight residual", x = "Year", fill = "Stratum")
#print(p2)
pltName <- paste( EBS.speciesnames[i],"plot", sep = '' )
myplot[[pltName]]<-p2}

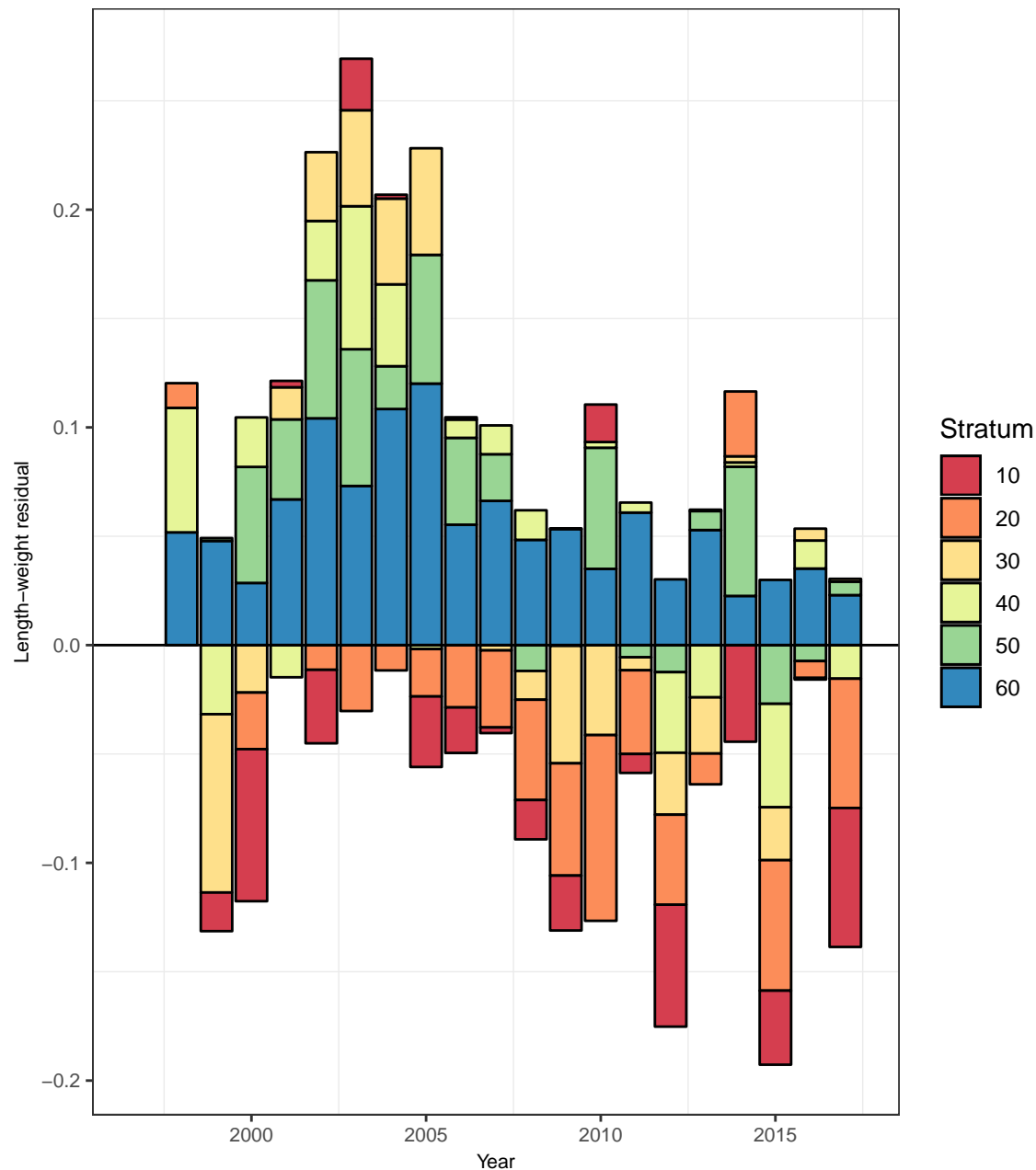
png("bystrata.png",width=6,height=7,units="in",res=300)
grid.arrange(grobs=myplot,ncol=1)
dev.off()

## pdf
## 2

grid.arrange(grobs=myplot,ncol=1)

```

## Pacific cod



```
write.csv(lwdata_by_strata,"lwdata_by_strata.csv",row.names=FALSE)
```

Here are the plots are by strata and year with individual plots for each strata (for J. Ianelli). Plots and data are output at the end.

```
lwdata_by_strata<-array(dim=c(0,7))
colnames(lwdata_by_strata)<-c("species","strata","yrs","ymean","yn","ysd","yse")
myplot<-list()
#By stratum graphs
for(i in 1:length(EBS.species)){
  tempdata<-subset(EBS.lwdata,EBS.lwdata$SPECIES_CODE==EBS.species[i])
  tempdata["residuals"]<-lw.resids(tempdata$LENGTH,tempdata$WEIGHT)
```

```

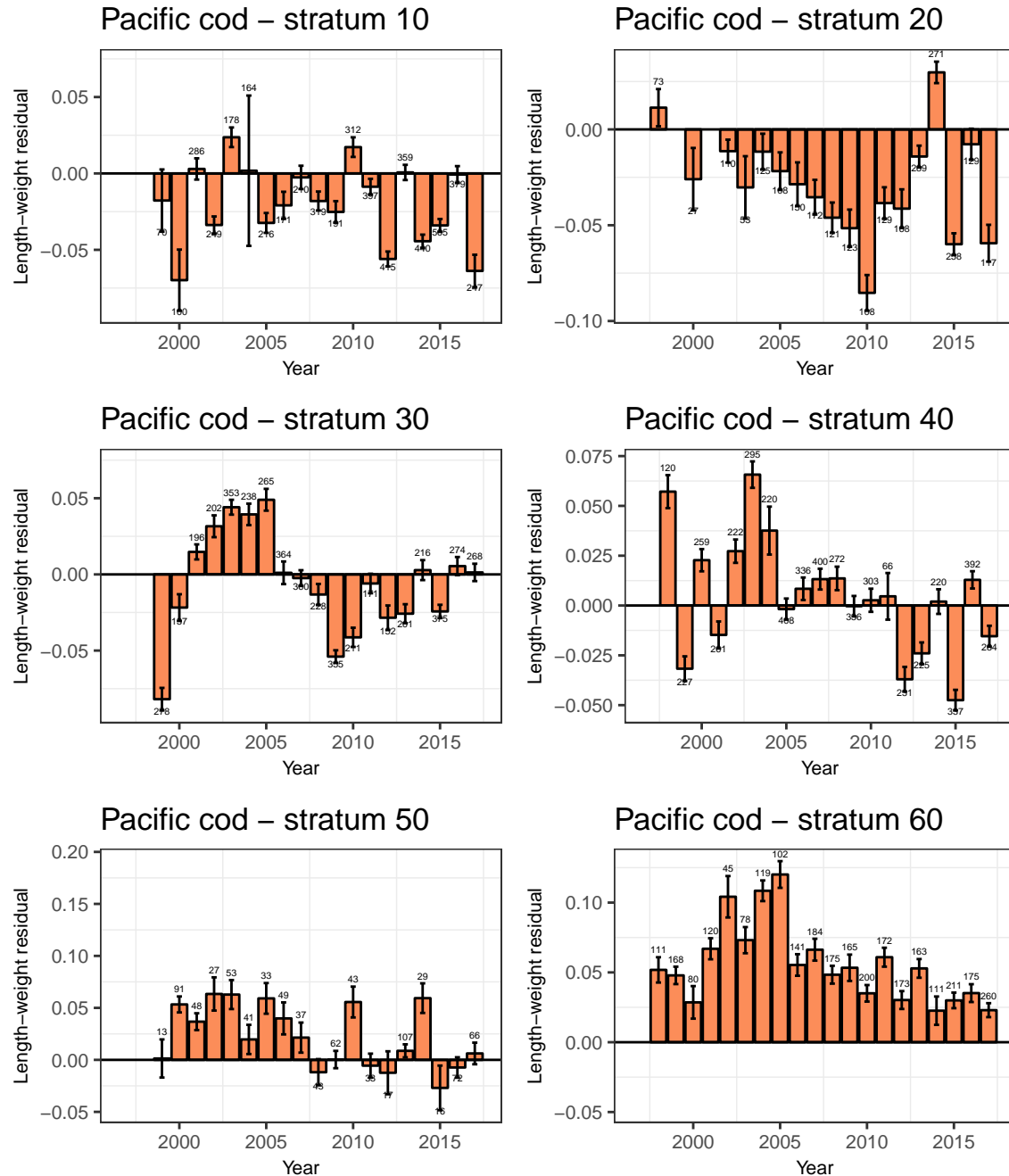
ymean=aggregate(tempdata$residuals,by=list(tempdata$YEAR,tempdata$STRATUM),mean)
ysd=aggregate(tempdata$residuals,by=list(tempdata$YEAR,tempdata$STRATUM),sd)
yn=aggregate(tempdata$residuals,by=list(tempdata$YEAR,tempdata$STRATUM),length)
yse=ysd$x/sqrt(yn$x)
data.summary<-data.frame(species=EBS.species[i],strata=ymean$Group.2,yrs=ymean$Group.1,ymean=ymean$ymean,ysd=ysd$ysd)
lwdata_by_strata<-rbind(lwdata_by_strata,data.summary)

ustrata<-unique(data.summary$strata)
for(j in 1:length(ustrata)){
  dat1<-subset(data.summary,data.summary$strata==ustrata[j])
  signymean<-sign(dat1$ymean)*(-.6)
  p2<-ggplot(data=dat1,aes(x=yrs,y=ymean,fill=factor(strata))) +
    geom_bar(position=position_dodge(),stat = "identity",col="black") +
    # geom_bar(position = position_dodge(), stat="identity", fill="cornflowerblue",col="black") +
    geom_errorbar(aes(ymin=ymean-yse, ymax=ymean+yse),width=0.30)+
    geom_text(data=dat1,aes(x=yrs,y=sign(ymean)*(abs(ymean)+yse),label=yn),vjust=signymean,size=1.5) +
    # geom_bar(data = dat2, aes(x=yrs, y=ymean, fill=factor(strata)),stat = "identity",col="black")+
    scale_fill_brewer(palette = "Spectral")+
    geom_hline(yintercept=0, color="black")+
    xlim(1996.5, 2017.5)+
    theme_bw() +
    theme(axis.text.x = element_text(size=8))+
    theme(axis.text.y = element_text(size=8))+
    theme(axis.title.x = element_text(size=8))+
    theme(axis.title.y = element_text(size=8))+
    theme(panel.grid.major = element_blank())+theme(legend.position="none")+
    labs(title = paste(EBS.speciesnames[i]," - stratum ",ustrata[j],sep=""), y = "Length-weight residual",
  pltName <- paste( EBS.speciesnames[i],ustrata[j],"plot", sep = ' ' )
  myplot[[pltName]]<-p2}

png(paste(EBS.speciesnames[i],"bystrata.png",sep="_"),width=6,height=7,units="in",res=300)
grid.arrange(grobs=myplot,ncol=2)
dev.off()
grid.arrange(grobs=myplot,ncol=2)}

```





```
write.csv(lwdata_by_strata,"lwdata_by_strata.csv",row.names=FALSE)
```

## GROUNDFISH CONDITION INDICES FOR ALASKA ESR

There are 3 r markdown scripts included in this package that can be used to produce the groundfish condition indices for the annual updates to the ecosystem contributions section of the SAFE documents. They are:

GOA\_GroundfishCondition.rmd

EBS\_GroundfishCondition.rmd

AI\_GroundfishCondition.rmd

These rmarkdown scripts are based on E. Siddon's conversion of the original R script and will create a word document and associated figures. It should only be necessary to update the length-weight data from RACEBASE and then the r markdown text (to reflect any new trends or results).