If you wish to conduct the same operation on a bunch of items use the For Loop

for filename in \*.cc.txt

do sort -k2 -b -n $filename > $filename.txt

cat $filename.txt > $filename

rm $filename.txt

done

Example: If you wish to ensure that the data for all individual country files (COUNTRYNAME.cc.txt) are sorted in descending order by date

for filename in \*.cc.txt

do sort -k2 -b -n $filename > $filename.txt

cat $filename.txt > $filename

rm $filename.txt

done

Notice that -k2 indicates the 2nd column which holds the year.

Look up help for sort by typing 'sort --help' (PCs) or 'man sort' (Mac)

Scripts

Generate scripts tha tyou can execute to run workflows of commands to achieve a task. Examples below start from hte gapminderdata.txt file

Type nano chk.sh

## two pound signs indicate comment lines that will nto be executed during the command run

## A script to automate orientation tasks

## Show current directory

pwd

## List directory content

ls -l

nano chk.sh

Now with some decoration

## Show current directory

echo "

You are in directory:

-----------------------------------------------"

pwd

## List directory content

echo "-----------------------------------------------

which contains these items:

-----------------------------------------------"

ls -l

echo "-----------------------------------------------

"

If you wish to execute scripts in ~/SWC\_Workshop/Scripts from anywhere under that's located under ~

export PATH=$PATH:~/SWC\_Workshop/Scripts

echo $PATH

Gapminder Data Scripts

Identify the ten countries with the highest gdpPercap and the lowes gdpPercap in 2007\_data

1) Get all entries for 2007 into one file 2007\_data.text

cp gapminderdata.txt ~/SWC\_Workshop/Analyses

cd ~/SWC\_Workshop/Analyses

grep --help

grep --help > help\_grep.txt

grep "\b2007\b" gapminderdata.txt > 2007\_data.txt

wc -l 2007\_data.txt

( grep "2007" gapminderdata.txt > 2007\_data.txt)

( wc -l 2007\_data.txt --> 144)

( less 2007\_data.txt)

## DoNotAddColumnHeaders!!!!!

2) Prepare the file 2007\_data.txt for sorting

Use' sed' to replace empty spaces in country names with '\_'; 2007\_data.txt is input as well as output

sed --help > help\_sed.txt

sed -i 's/ /\_/g' 2007\_data.txt

3) Extract the countries with the 10 highest and the 10 lowest GDPs in 2007

sort -k6 -b -n 2007\_data.txt > 2007\_sortedbyGDP.txt

tail -10 2007\_sortedbyGDP.txt | sort -k6bnr > 10\_richest\_countries\_allData.txt

head -10 2007\_sortedbyGDP.txt > 10\_poorest\_countries\_allData.txt

ls -l

rm -I \*

Write a script that does the job for you - one last time copying and pasting and we are done

nano PoorCountryRichCountry.sh

sort -k6 -b -n 2007\_data.txt > 2007\_sortedbyGDP.txt

head -10 2007\_sortedbyGDP.txt > 10\_poorest\_countries.txt

tail -10 2007\_sortedbyGDP.txt > 10\_richest\_countries.txt

cat 10\_poorest\_countries.txt 10\_richest\_countries.txt > PoorCountryRichCountry.txt

Save

sh PoorCountryRichCountry.sh

cat PoorCountryRichCountry.txt

(mv -i PoorCountryRichCountry.sh ~/SWC\_Workshop/scripts/)

Modify the script to put the rich countries on top and give you everything in descending rankings

nano PoorCountryRichCountry.sh

change 2nd row to

head -10 2007\_sortedbyGDP.txt | sort -k6bnr > 10\_poorest\_countries.txt

change 3nd row to

tail -10 2007\_sortedbyGDP.txt | sort -k6bnr | cut -f1 > 10\_richest\_countries.txt

change 4th row to

cat 10\_richest\_countries.txt 10\_poorest\_countries.txt > RichCountryPoorCountry.txt

Save as RichCountryPoorCountry.sh

sh RichCountryPoorCountry.sh

cat RichCountryPoorCountry.txt

Modify RichCountryPoorCountry.sh to only give you the countries and the GDP figures, sans decimals

nano RichCountryPoorCountry.sh

change 2n row to

head -10 2007\_sortedbyGDP.txt | sort -k6bnr | cut -f1,6 | cut -d . -f 1 > 10\_poorest\_countries.txt

change 3rd row to

tail -10 2007\_sortedbyGDP.txt | sort -k6bnr | cut -f1,6 | cut -d . -f 1 > 10\_richest\_countries.txt

change 4th row to

cat 10\_richest\_countries.txt 10\_poorest\_countries.txt > RichCountryPoorCountry\_Countries.txt

Save as RichCountryPoorCountry\_Countries.sh

sh RichCountryPoorCountry\_Countries.sh

cat RichCountryPoorCountry\_Countries.txt

Modify PoorCountryRichCountry.sh to give you the continents without repetition in ascending order

nano PoorCountryRichCountry.sh

change 2nd row to

head -10 2007\_sortedbyGDP.txt sort | -k4b | cut -f4 | uniq > poorest\_continents.txt

change 3rd row to

tail -10 2007\_sortedbyGDP.txt | sort -k4b | cut -f4 | uniq > richest\_continents.txt

Change 4th row to

cat richest\_continents.txt poorest\_continents.txt > PoorCountryRichCountry\_Continents.txt

Save as PoorContinentsRichContinents.sh

type sh PoorContinentsRichContinents.sh

cat PoorCountryRichCountry\_Continents.txt

On Your Own: Modify the script to provide you with the longest and least life expectancy