**IT497 Exam *#1***

*For this exam, create an R notebook. Answer question 1-3 in the notebook. Knit the notebook into either a Word Document or HTML file. Submit both the \*.Rmd file and either the Word Document or HTML file knitted from the R notebook. For question 4, create a video describing how you solved it. The video may by a screen cast or shot from your phone or camera. For question 4, upload the video and any other file needed to support your answer.*

# Each student has different assignment for questions 1 and 2

Here are the assignments (questions are below)

|  |  |  |
| --- | --- | --- |
| **Question 1** | | **Question 2 Peak Name** |
| **Full Name** | **Region** |
| Agyemang, Eric | 1 | Everest |

**Answer each question fully – use the table above to determine the proper states/peaks.**

**Question *#1***

**Step 1:**

Install the remotes package

install.packages("remotes") (you only need to do this one time)

**Step 2:**

Go to <https://jrosen48.github.io/tidykids/index.html>and install the **tidykids** package (you only need to do this one time)

The instructions are on the website. Also, the data is described on the website. You will use this data for your exam.

**Step 3.** Use the **tidykids** data to answer your assigned questions.

|  |  |
| --- | --- |
| **State** | **Region** |
| Alabama | 1 |
| Florida | 1 |
| Georgia | 1 |
| Louisiana | 1 |
| Mississippi | 1 |
| Tennessee | 1 |
| Arkansas | 1 |
| Texas | 1 |
| Montana | 2 |
| North Dakota | 2 |
| South Dakota | 2 |
| Wyoming | 2 |
| New Mexico | 2 |
| Idaho | 2 |
| Kansas | 2 |
| Nebraska | 2 |
| Oklahoma | 2 |
| Utah | 3 |
| Colorado | 3 |
| Oregon | 3 |
| Washington | 3 |
| Alaska | 3 |
| Arizona | 3 |
| California | 3 |
| Hawaii | 3 |
| Nevada | 3 |
| South  Carolina | 4 |
| Kentucky | 4 |
| North  Carolina | 4 |
| Virginia | 4 |
| West Virginia | 4 |
| Maryland | 4 |
| Delaware | 4 |
| Pennsylvania | 4 |
| Connecticut | 5 |
| Maine | 5 |
| Massachusetts | 5 |
| New  Hampshire | 5 |
| New Jersey | 5 |
| New York | 5 |
| Rhode Island | 5 |
| Vermont | 5 |
| Illinois | 6 |
| Indiana | 6 |
| Iowa | 6 |
| Michigan | 6 |
| Minnesota | 6 |
| Ohio | 6 |
| Wisconsin | 6 |
| Missouri | 6 |

For each of the following, **use 2016 numbers only**

**Question 1 Output 1**. Find the 3 states that spend the most (raw) on PK12ed = Public spending on elementary and secondary education (K-12). Show only these 3 states. List the 3 states from highest to lowest **in** PK12ed spending.

**Question 1 Output 2**. Find the 3 states that spend the least (raw) on PK12ed. Show only these 3 states. List the 3 states from lowest to highest **in** PK12ed spending.

**Question 1 Output 3**. Find the 3 states that spend the most (adjusted per child) on SNAP. The Supplemental Nutrition Assistance Program (SNAP) "provides nutrition benefits to supplement the food budget of needy families so they can purchase healthy food and move towards selfsufficiency." Show only these 3 states. List the 3 states from highest to lowest **in** SNAP spending.

**Question 1 Output 4**. Find the 3 states that spend the least (adjusted per child) on SNAP. Show only these 3 states. List the 3 states from lowest to highest **in** SNAP spending.

**Question 1 Output 5**. Find the 3 states that have had the largest increase **in** (inflation adjusted) spending on Higher Ed spending from 1997 to 2016. Show only these 3 states. List the 3 states from highest to lowest **in** highered spending.

**Question *#2***

**Step 1.** Read in the data. The data comes from The Himalayan Database.

The Himalayan Database is a compilation of records for all expeditions that have climbed in the Nepal Himalaya. The data cover all expeditions from 1905 through Spring 2019

*# Note readr is part of tidyverse. There is no need to library(readr)*

expeditions <-

readr::read\_csv('https://raw.githubusercontent.com/rfordatascience/tid ytuesday/master/data/2020/2020-09-22/expeditions.csv')

Using **your assigned peak**, answer the following:

**Question 2 Output *#1*** *Which year had the most member deaths (for your assigned peak only). Show expedition name and year and member deaths.*

**Question 2 Output *#2*** *Which expedition had the most deaths (for your assigned peak only) Include both member and hired staff deaths combined. Show expedition name and year and total deaths.*

**Question 2 Output *#3*** *Which expedition had the highest percent of members die (for your assigned peak only). To find this, divide member deaths by members. Show expedition name and year and percent of member deaths.*

**Question 2 Output *#4*** *Which expedition had the highest percent of staff die (for your assigned peak only). To find this, divide hired staff deaths by hired staff. Show expedition name and year and percent of staff deaths.*

**Question 2 Output *#5*** *Which year had the most climbers successfully reach the top of your assigned peak? Hint: use* **filter(termination\_reason == 'Success (main peak)')**

**Question *#3***

**Step 1**. Use **rvest** to download the **Rank, Distribution, HPD** of the top 100 Linux distros from <https://distrowatch.com/>

**Question 3 Output *#1****. Show the Rank, Distribution, HPD of all Linux distros with a* ***HPD > 1000.***

**Question 3 Output *#2****. Show the Rank, Distribution, HPD of all* ***UBUNTU*** *distros.*

**Question *#4***

**Redo question *#3 NOT using rvest****. You must obtain the data using a different method that does not involve rvest in anyway. It may or may not involve R.*