

95-868 M23 Project

Due, Saturday, August 5, 2023 11:59 pm ET

Vera (Shuyao) Fan

shuyaof

```
# Import packages
library(tidyverse)
library(lubridate)
library(scales)
library(ggrepel)
library(dplyr)
```

Problem 1

Which regions have seen the largest *percent* increase in visitors (at national parks *only*) from 1985 to 2015?

Show this as a horizontal bar chart with the percent change on the x-axis and region on the y-axis. Ensure the bars are properly ordered. Show the regions with their full names instead of abbreviations (hint: Wikipedia will helpfully provide these, consider `case_when`) and include the total number of national parks in each region in parentheses as part of the axis labels. Regions with a positive percent change should have a different color from regions with a negative percent change.

```
# A tibble: 21,483 x 13
   year gnis_id geometry metadata number_of_records parkname region state
   <dbl>   <dbl> <chr>      <chr>          <dbl> <chr>      <chr> <chr>
1  1904 1163670 POLYGON      <NA>             1 Crater La~ PW    OR
2  1941 1531834 MULTIPOLYGON <NA>             1 Lake Roos~ PW    WA
3  1961 2055170 MULTIPOLYGON <NA>             1 Lewis and~ PW    WA
4  1935 1530459 MULTIPOLYGON <NA>             1 Olympic   PW    WA
5  1982 277263 POLYGON      <NA>             1 Santa Mon~ PW    CA
6  1919 578853 MULTIPOLYGON <NA>             1 <NA>      NE    ME
7  1969 1329499 MULTIPOLYGON <NA>             1 <NA>      IM    TX
```

```

8 1967 589056 POLYGON <NA> 1 <NA> NE MD
9 1944 1377082 POLYGON <NA> 1 <NA> IM TX
10 1989 302659 POLYGON <NA> 1 <NA> SE FL

```

```
# i 21,473 more rows
```

```
# i 5 more variables: unit_code <chr>, unit_name <chr>, unit_type <chr>,
# visitors <dbl>, region_full_names <chr>
```

```
# A tibble: 14 x 4
```

```
# Groups:   region, region_full_names [7]
  region region_full_names year visitors
  <chr>   <chr>             <dbl>   <dbl>
1 AK     Alaska             1985   1103491
2 AK     Alaska             2015   2116311
3 IM     Intermountain       1985  16475380
4 IM     Intermountain       2015  31259549
5 MW     Midwest             1985   4123360
6 MW     Midwest             2015   6141356
7 NC     National Capital     1985    458395
8 NC     National Capital     2015    407351
9 NE     Northeast            1985   5678655
10 NE     Northeast            2015   4133057
11 PW     Pacific West          1985  13941994
12 PW     Pacific West          2015  18738798
13 SE     Southeast            1985  12710739
14 SE     Southeast            2015  13461907
```

```
# A tibble: 7 x 5
```

```
# Groups:   region, region_full_names [7]
  region region_full_names visitors_1985 visitors_2015 percent_change
  <chr>   <chr>             <dbl>         <dbl>         <dbl>
1 AK     Alaska             1103491       2116311         0.918
2 IM     Intermountain       16475380      31259549         0.897
3 MW     Midwest             4123360       6141356         0.489
4 NC     National Capital     458395        407351        -0.111
5 NE     Northeast            5678655      4133057        -0.272
6 PW     Pacific West        13941994      18738798         0.344
7 SE     Southeast           12710739      13461907         0.0591
```

```
# A tibble: 7 x 2
```

```
  region distinct_parks
  <chr>         <int>
1 AK             9
```

```

2 IM          18
3 MW          7
4 NC          1
5 NE          2
6 PW         17
7 SE          7

```

```
# A tibble: 7 x 6
```

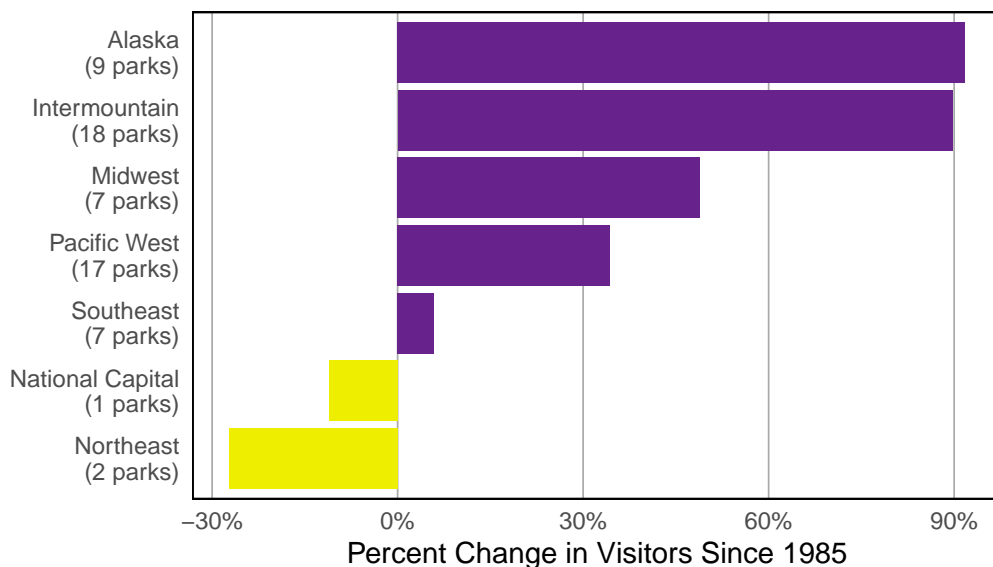
```
# Groups:   region, region_full_names [7]
```

```

  region region_full_names visitors_1985 visitors_2015 percent_change
  <chr>   <chr>             <dbl>         <dbl>         <dbl>
1 AK     Alaska             1103491        2116311         0.918
2 IM     Intermountain      16475380       31259549         0.897
3 MW     Midwest            4123360        6141356         0.489
4 NC     National Capital    458395         407351        -0.111
5 NE     Northeast           5678655        4133057        -0.272
6 PW     Pacific West       13941994       18738798         0.344
7 SE     Southeast          12710739       13461907         0.0591
# i 1 more variable: distinct_parks <int>

```

National parks in the Alaska & Intermountain regions saw the largest increase in visitors from 1985 to 2015



Explanation:

During the period between 1985 and 2015, the Alaska and Intermountain regions emerged as the standout stars, experiencing a remarkable surge in visitor percentage. Despite having only

half the number of parks as the Intermountain region, Alaska's allure proved irresistible to travelers. In contrast, the Midwest, Pacific, and Southeast regions received relatively modest increases in visitors, each falling short of a 60% rise over the same 30-year span. However, not all regions were fortunate enough to witness a positive trend. Both the National Capital and Northeast regions faced a decline in the number of visitors. It's plausible that the limited number of parks in these areas contributed to their relative unpopularity.

Problem 2

Which type of entity is most frequented?

The National Park Service supports many different park types: national parks, national monuments, etc. Determine the four most common types of park (called `unit_type` in the data set) by counting the number of distinct occurrences in the data set. Create a faceted line graph (`facet_wrap`) for these four park types, with time from 1985 to 2015 on the x-axis and visitors per 100 people in the US for each year on the y-axis. Make sure that the faceted plots are ordered by the highest value in 2015. (Hint: you can do this by manually setting the levels of the `unit_type` variable using `factor` as part of your data manipulation steps.) Change the style of the facet labels/headers in some way.

```
# A tibble: 4 x 2
  unit_type      n
  <chr>      <int>
1 National Historic Site    78
2 National Monument       75
3 National Park            61
4 National Historical Park   42
```

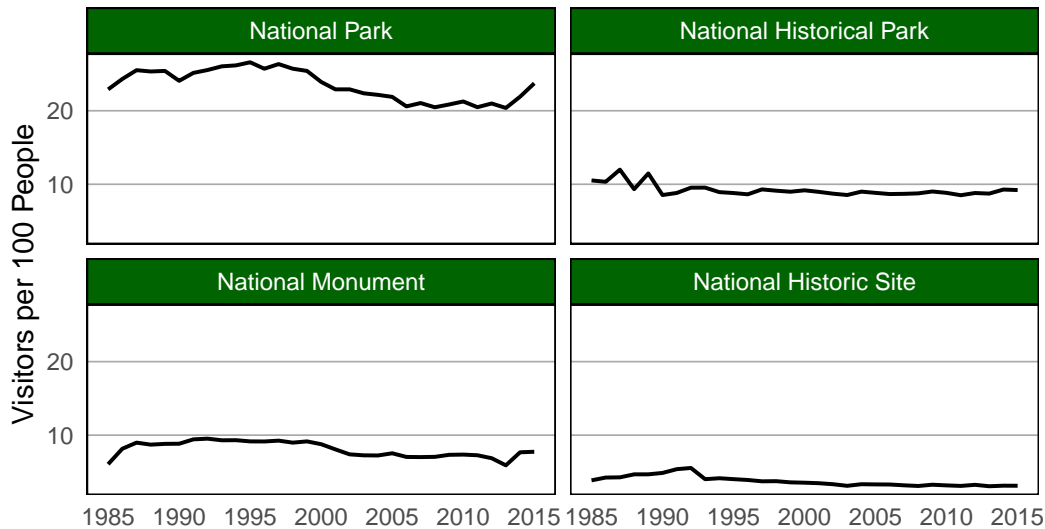
```
# A tibble: 31 x 2
  year population
  <dbl>      <dbl>
1  1985  237923796
2  1986  240132883
3  1987  242288918
4  1988  244498983
5  1989  246819230
6  1990  249464396
7  1991  252153092
8  1992  255029699
9  1993  257782608
10 1994  260327021
# i 21 more rows
```

```
# A tibble: 124 x 3
# Groups:   unit_type [4]
  unit_type      year visitors
  <chr>         <dbl>    <dbl>
1 National Historic Site 1985  9169947
2 National Historic Site 1986 10174532
3 National Historic Site 1987 10330165
4 National Historic Site 1988 11423594
5 National Historic Site 1989 11539696
6 National Historic Site 1990 12130189
7 National Historic Site 1991 13549889
8 National Historic Site 1992 14122360
9 National Historic Site 1993 10352598
10 National Historic Site 1994 10782297
# i 114 more rows
```

```
# A tibble: 124 x 5
# Groups:   unit_type [4]
  unit_type      year visitors population percentage
  <chr>         <dbl>    <dbl>    <dbl>    <dbl>
1 National Historic Site 1985  9169947 237923796 3.85
2 National Historic Site 1986 10174532 240132883 4.24
3 National Historic Site 1987 10330165 242288918 4.26
4 National Historic Site 1988 11423594 244498983 4.67
5 National Historic Site 1989 11539696 246819230 4.68
6 National Historic Site 1990 12130189 249464396 4.86
7 National Historic Site 1991 13549889 252153092 5.37
8 National Historic Site 1992 14122360 255029699 5.54
9 National Historic Site 1993 10352598 257782608 4.02
10 National Historic Site 1994 10782297 260327021 4.14
# i 114 more rows
```

```
# A tibble: 4 x 3
# Groups:   unit_type [4]
  unit_type      year visitors
  <chr>         <dbl>    <dbl>
1 National Park      2015 76258329
2 National Historical Park 2015 29615999
3 National Monument   2015 24888632
4 National Historic Site 2015 10027951
```

National parks remain the most frequently-visited entity supported by NPS



Explanation:

Over the 30-year span, National Parks have consistently remained the most beloved type of visited entities, attracting over 20 visitors per 100 people. Meanwhile, National Historic Sites have been less popular, with less than 5 visitors per 100 people. National Historical Parks and National Monuments have shared a similar level of popularity, experiencing a stable number of visitors over time. However, there has been a noticeable decline in the number of visitors to both National Parks and National Monuments a little before the year 2000.

Problem 3

How do gas prices affect total visitors?

Create a scatter plot showing the gas price against visits per 100 people in the US, from 1980 onward for national parks only in the Southeast and Pacific-West regions. The y-axis should be gas prices in 2015 dollars and the x-axis should be visitors per 100 people in the US for each year. (Hint: you will have to `left_join` with the `state_pop` data as well as the `gas_price` data.) Add a line of best-fit for each group and instead of using a legend, label a point from each group.

```
# A tibble: 74 x 3
# Groups:   year [37]
  year region visitors
```

```

      <dbl> <chr>      <dbl>
1  1980 PW      13307431
2  1980 SE      11485617
3  1981 PW      14296374
4  1981 SE      11164102
5  1982 PW      14423689
6  1982 SE      11287112
7  1983 PW      14560368
8  1983 SE      11568303
9  1984 PW      15259647
10 1984 SE      11657400
# i 64 more rows

```

```

# A tibble: 36 x 2
  year population
  <dbl>      <dbl>
1  1980  227224719
2  1981  229465717
3  1982  231664452
4  1983  233791992
5  1984  235824901
6  1985  237923796
7  1986  240132883
8  1987  242288918
9  1988  244498983
10 1989  246819230
# i 26 more rows

```

```

# A tibble: 74 x 5
# Groups:   year [37]
  year region visitors population visitors_per_100
  <dbl> <chr>      <dbl>      <dbl>      <dbl>
1  1980 PW      13307431  227224719      5.86
2  1980 SE      11485617  227224719      5.05
3  1981 PW      14296374  229465717      6.23
4  1981 SE      11164102  229465717      4.87
5  1982 PW      14423689  231664452      6.23
6  1982 SE      11287112  231664452      4.87
7  1983 PW      14560368  233791992      6.23
8  1983 SE      11568303  233791992      4.95
9  1984 PW      15259647  235824901      6.47
10 1984 SE      11657400  235824901      4.94

```

```
# i 64 more rows
```

```
# A tibble: 74 x 7
```

```
# Groups:   year [37]
```

	year	region	visitors	population	visitors_per_100	gas_current	gas_constant
	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1980	PW	13307431	227224719	5.86	1.19	2.95
2	1980	SE	11485617	227224719	5.05	1.19	2.95
3	1981	PW	14296374	229465717	6.23	1.31	2.97
4	1981	SE	11164102	229465717	4.87	1.31	2.97
5	1982	PW	14423689	231664452	6.23	1.22	2.6
6	1982	SE	11287112	231664452	4.87	1.22	2.6
7	1983	PW	14560368	233791992	6.23	1.16	2.37
8	1983	SE	11568303	233791992	4.95	1.16	2.37
9	1984	PW	15259647	235824901	6.47	1.13	2.23
10	1984	SE	11657400	235824901	4.94	1.13	2.23

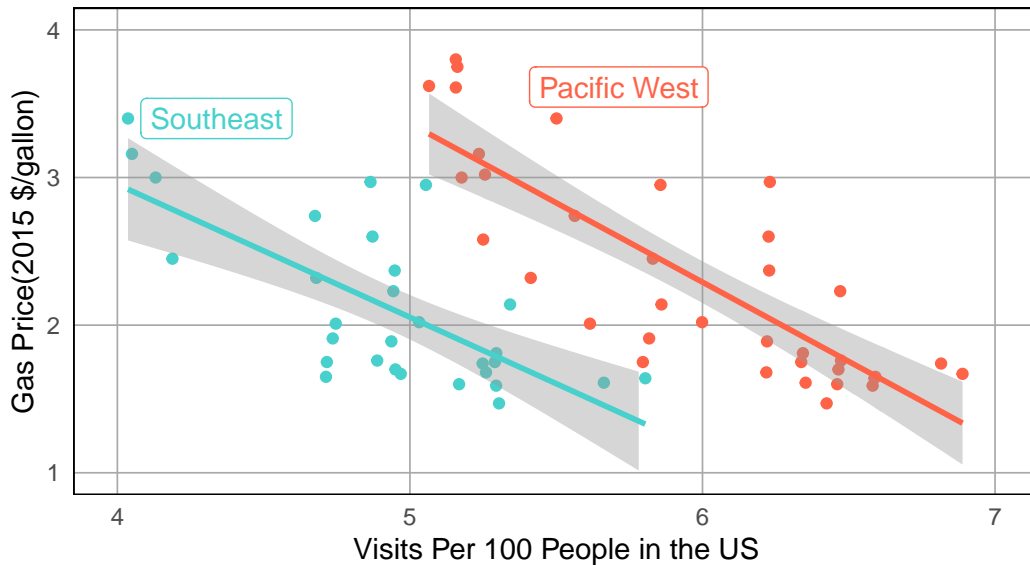
```
# i 64 more rows
```

```
# A tibble: 2 x 7
```

```
# Groups:   year [1]
```

	year	region	visitors	population	visitors_per_100	gas_current	gas_constant
	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2014	PW	17540230	318857056	5.50	3.37	3.4
2	2014	SE	12870467	318857056	4.04	3.37	3.4

Visitors to national parks in the Pacific West and Southeast region seem to be negatively affected by gas prices



Explanation:

Based on the scatter plot, a clear pattern emerges, indicating an inverse relationship between visits to national parks and gas prices in both the Pacific West and Southeast regions. Notably, the ratio between gas price and visits per 100 people appears to align similarly for both regions. Additionally, it's interesting to observe that while the Pacific West region experiences higher gas prices, when number of visits remains comparable to the Southeast region.

Problem 4

Which national parks have become the most popular?

Create a line graph with one line per park, for national parks only, and time on the x-axis, 1985 to 2015. The value on the y-axis should be visitors per 100 people in the US for each year. (Hint: you will have to `left_join` with the `state_pop` data.) Figure out which two parks are consistently the most visited and highlight those lines. Use some kind of annotation method to label those two lines with the name of the park.

```
# A tibble: 1,871 x 3
# Groups:   year [31]
   year unit_name      visitors
  <dbl> <chr>         <dbl>
1  1985 Acadia National Park 3745570
```

```

2 1985 Arches National Park 363464
3 1985 Badlands National Park 950242
4 1985 Big Bend National Park 188045
5 1985 Biscayne National Park 460921
6 1985 Black Canyon of the Gunnison National Park 266012
7 1985 Bryce Canyon National Park 500782
8 1985 Canyonlands National Park 116672
9 1985 Capitol Reef National Park 320503
10 1985 Carlsbad Caverns National Park 732482
# i 1,861 more rows

```

```

# A tibble: 31 x 2
  year population
  <dbl>      <dbl>
1 1985 237923796
2 1986 240132883
3 1987 242288918
4 1988 244498983
5 1989 246819230
6 1990 249464396
7 1991 252153092
8 1992 255029699
9 1993 257782608
10 1994 260327021
# i 21 more rows

```

```

# A tibble: 1,871 x 5

```

```

# Groups:   unit_name [61]

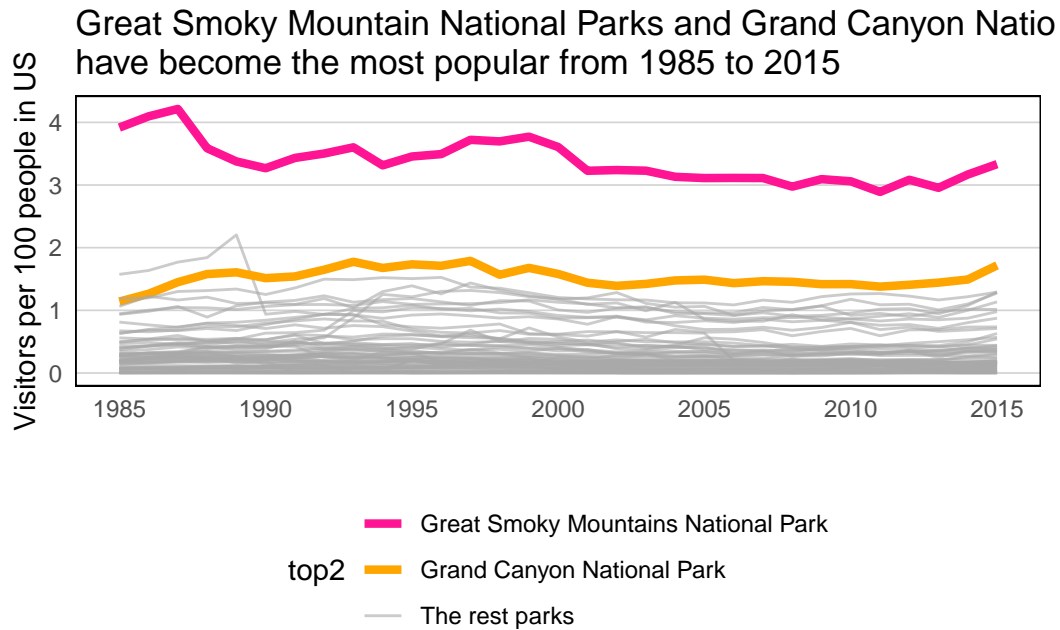
```

	unit_name	year	visitors	population	visitors_per_100
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Great Smoky Mountains National Pa~	1987	10209841	242288918	4.21
2	Great Smoky Mountains National Pa~	1986	9836306	240132883	4.10
3	Great Smoky Mountains National Pa~	1985	9319290	237923796	3.92
4	Great Smoky Mountains National Pa~	1999	10283598	272690813	3.77
5	Great Smoky Mountains National Pa~	1997	9965075	267783607	3.72
6	Great Smoky Mountains National Pa~	1998	9989395	270248003	3.70
7	Great Smoky Mountains National Pa~	2000	10175812	282162411	3.61
8	Great Smoky Mountains National Pa~	1993	9283848	257782608	3.60
9	Great Smoky Mountains National Pa~	1988	8770781	244498983	3.59
10	Great Smoky Mountains National Pa~	1992	8931690	255029699	3.50

```

# i 1,861 more rows

```



Explanation:

The Great Smoky Mountains National Park stands out as a favorite among visitors, attracting over three visitors for every 100 people. Its popularity reached a peak around 1987. Although there has been a slight decline in recent years, it remains a cherished destination for many. Following closely behind is the Grand Canyon National Park, which has more than one visitor per 100 people. This park maintains a steady level of popularity compared to other national parks. Most of the remaining national parks, in contrast, receive less than one visitor per 100 people.

Problem 5

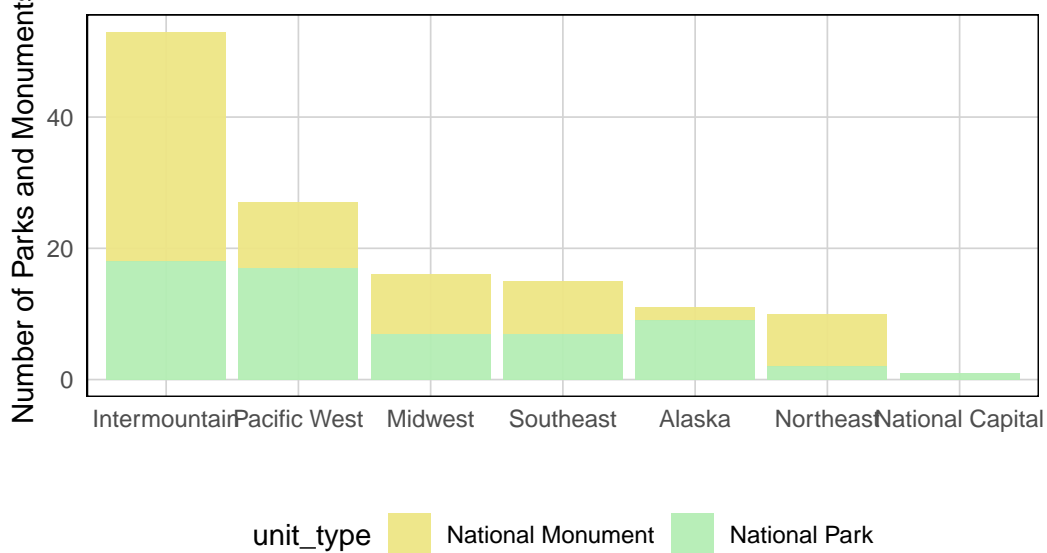
Which region has the most national parks and national monuments?

Using the data set(s) available, formulate your own question and answer it using any chart type that we've discussed in class. In addition to the summary of your conclusions, discuss why you chose the particular chart type that you did for the question you came up with.

```
# A tibble: 13 x 3
# Groups:   region_full_names [7]
  region_full_names unit_type      count
  <chr>             <chr>         <int>
1 Alaska           National Monument      2
```

2	Alaska	National Park	9
3	Intermountain	National Monument	35
4	Intermountain	National Park	18
5	Midwest	National Monument	9
6	Midwest	National Park	7
7	National Capital	National Park	1
8	Northeast	National Monument	8
9	Northeast	National Park	2
10	Pacific West	National Monument	10
11	Pacific West	National Park	17
12	Southeast	National Monument	8
13	Southeast	National Park	7

Intermountain region has the most of National Parks and National Monuments from 1985 to 2015



Explanation:

The bar plot displays a unique count of parks and monuments per region from the years 1985 to 2015. The Intermountain region stands out with the highest number of national parks and monuments, offering travelers a diverse array of destinations to explore. Conversely, the National Capital region doesn't have any national monuments and only comprises a limited number of parks. The Pacific West Region and Alaska regions are the second and third regions with the highest number of national parks.

I specifically chose a stacked bar chart, setting it apart from the previous four questions. In order to enhance visual clarity, I focused on just two unit types for comparison. Therefore,

the data is presented in a more easily digestible manner.

As a traveler planning to visit national parks and monuments, this chart would undoubtedly lead me to favor the Intermountain region as my destination of choice. Its abundant selection of parks and monuments provides a thrilling range of options to choose from.