Problem Set 3

Bern DySart

2024-02-14

1: See table below

```
water1.0 =
  water |>
  select(
    countryModule,
    interviewYear,
    waterTimeMins
  )

water2.0 =
  water1.0 |>
  group_by(countryModule) |>
  summarise(
    n = n(),
    year = list(unique(interviewYear)),
    wtdmean = mean(waterTimeMins, na.rm = TRUE),
    wtdsd = sd(waterTimeMins, na.rm = TRUE),
    wtdmed = median(waterTimeMins, na.rm = TRUE),
    perplot = mean(waterTimeMins == 0, na.rm = TRUE) * 100
  )

kable(water2.0, digits = 0)
```

countryModule	n	year	wtdmean	wtdsd	wtdmed	perplot
BF6	28849	2010	20	19	15	9
BJ6	32486	2012, 2011	14	31	5	34
CD6	33656	2014, 2013	33	31	30	6
CI6	16298	2012, 2011	15	29	5	43
CM6	24163	2011	20	27	10	17
ET6	27715	2003	53	72	30	13
GA6	12854	2012, NA	21	36	5	45
GH5	15697	2008	18	26	10	16
GN6	16380	2012	23	25	20	23
KE5	13105	2009, 2008, NA	28	41	15	28
KM6	8141	2012	10	25	0	69
LB6	16229	2013	17	19	10	8
LS5	14300	2009, 2010	20	24	10	13
MD5	31744	2008, 2009	17	41	10	15
ML6	20598	2012, 2013	10	28	3	37

countryModule	n	year	wtdmean	wtdsd	wtdmed	perplot
MW5	43732	2010	30	32	20	9
MZ6	22075	2011	29	50	15	24
NG6	58812	2013	20	29	10	21
NM6	12171	2013, NA	14	26	0	50
RW6	19246	2011, 2010	36	33	30	6
SL6	25749	2013	21	23	15	10
SN6	26564	2011, 2010	12	31	0	54
SZ5	8025	2006, 2007, NA	21	31	10	32
TG6	16280	2013, 2014	24	29	15	13
TZ5	17251	2010, 2009	30	39	16	13
UG6	16740	2011	46	50	30	10
ZM6	30871	2013, 2014	18	24	10	24
ZW6	14195	2010, 2011, NA	19	30	10	35

2. See summary data below

```
Exp11.fixed =
  Exp11|>
  rename(
    Bonus = Pay
  ) |>
  select (start, CompCode, Bonus)

Exp10.fixed =
  Exp10 |>
  select(start, CompCode, Bonus)

Exp1011 =
  full_join(
    Exp11.fixed,
    Exp10.fixed,
    by = join_by(CompCode)
  )

full =
  inner_join(
    Exp1011,
    Crid,
    by = join_by(CompCode)
  )

full2.0 =
  full |>
  select(AssignmentID, rID, CompCode, Bonus.x)

str(full2.0)
```

```
## tibble [1,301 x 4] (S3: tbl_df/tbl/data.frame)
##  $ AssignmentID: chr [1:1301] "3D4CH1LGEEYYCG644RU9PW5ZNQ09GN" "3HL8HNGX496SYGA9T7OR7B5NMB09FO" "3E7
##  $ rID         : chr [1:1301] "A22OPXKTGLCX7B" "A2N4Q6OTCBWBL4" "A22HIX1M4QXZBB" "A27X2HWYZ76M5K" ..
##  $ CompCode    : chr [1:1301] "R_2cqk5ElS6J1eXWu" "R_3G95bZjmyBO2OFV" "R_bpwQBDMhEyguBTX" "R_3m4vguJ
##  $ Bonus.x     : num [1:1301] 65 64 63 65 65 62 61 62 63 63 ...
```

3. See Table Below

```
wordle1.0 =
  wordle |>
  pivot_longer(
    cols = starts_with('l'),
    names_to = 'id',
    values_to = 'letters'
  )

wordle2.0 =
    wordle1.0 |>
    count(letters) |>
    arrange(desc(n))

kable(wordle2.0)
```

letters	n
s	6665
e	6662
a	5990
o	4438
r	4158
i	3759
l	3371
t	3295
n	2952
u	2511
d	2453
y	2074
c	2028
p	2019
m	1976
h	1760
g	1644
b	1627
k	1505
f	1115
w	1039
v	694
z	434
j	291
x	288
q	112

4. See table below for highest and lowest scoring words

```
wordle3.0 =
  left_join(
    wordle1.0,
    wordle2.0,
    by = join_by(letters)
```

```
  )

wordle4.0 =
  wordle3.0 |>
  group_by(word) |>
  summarise(
    n = sum(n)
  ) |>
  filter(n == max(n) | n == min(n)) |>
  arrange(desc(n))

kable(wordle4.0)
```

word	n
esses	33319
fuzzy	6568

5. See table for highest and lowest scoring words without repeating letters

```
wordle5.0 =
  wordle3.0 |>
  group_by(word) |>
  filter(dups == FALSE) |>
  summarise(
    n = sum(n)
  ) |>
  filter(n == max(n) | n == min(n)) |>
  arrange(desc(n))

kable(wordle5.0)
```

word	n
aeros	27913
arose	27913
soare	27913
jumby	8479