

Olimpia Pozauć, Oliwia Krzemińska

# Analiza czynników wpływających na sukces wypraw w Himalaje

16 grudnia 2021

## 1. Wprowadzenie i opis danych

Raport stanowi analizę danych rzeczywistych dotyczących członków ekspedycji i wypraw wspinaczkowych, które odbyły się w Himalajach. Dane pochodzą ze strony [kaggle](#). Zgodnie z informacjami na stronie, zostały one opracowane na podstawie archiwów dziennikarki Elizabeth Hawley, uzupełnione informacjami z książek, dzienników i korespondencji z himalaistami. Dane obejmują łącznie wszystkie wyprawy na przestrzeni lat 1905-2019.

W raporcie analizowany jest jeden z trzech plików dostępnych na stronie. Plik `members.csv` zawiera, między innymi, dane dotyczące roku ekspedycji, nazwy szczytów, informacje o pochodzeniu, płci oraz wieku członków wyprawy.

Sukces dla każdego członka wyprawy rozumiemy jako zdobycie szczytu, a celem naszej analizy jest zbadanie jakie czynniki i w jakim stopniu wpływają na odniesienie sukcesu.

## 2. Analiza zmiennych

Przeprowadzona analiza dotyczy tylko jednego, najczęściej wybieranego szczytu *Everest*, dla którego jest aż 21813 obserwacji. Dla tego szczytu pierwsze dane pochodzą z 1921, a ostatnie z 2019 roku.

Ponieważ jako sukces uznajemy w naszym raporcie zdobycie szczytu, analizujemy tylko członków, którym przypisano podczas ekspedycji rolę `'Climber'`. Dla pozostałych członków zespołu, mimo wielkiego wsparcia, celem samym w sobie nie jest prawdopodobnie osiągnięcie wierzchołka góry. Mamy wówczas 10797 obserwacji.

Dla przejrzystości usuwamy kolumny, które nie są istotne dla przedstawionego problemu badawczego, są to `expedition_id`, `member_id`, `peak_id`, `peak_name`, `hired`, `solo`, `died`, `death_cause`, `death_height_metres`, `injured`, `injury_type`, `injury_height_metres`.

W raporcie analizujemy następujące zmienne:

- `year` - rok ekspedycji,
- `season` - pora roku, w której odbyła się wspinaczka,
- `sex` - płeć uczestników,
- `age` - wiek uczestników,
- `expedition_role` - rola, jaką dany uczestnik ma w grupie,
- `highpoint_metres` - najwyższa wysokość, którą osiąga dany uczestnik,
- `success` - sukces, czyli zdobycie szczytu,

— `oxygen_used` - użycie tlenu.

Jako zmienne ciągłe uznajemy `highpoint_meters`, czyli najwyższą wysokość, jaką udało się osiągnąć danemu himalaistcie oraz zmienną `age` oznaczającą wiek wspinacza.

Pozostałe zmienne to zmienne katgoriczne.

### 3. Analiza graficzna

#### 3.1. Zależność najwyższej osiągananej wysokości od wieku uczestników

W pierwszej kolejności możemy sprawdzić, czy zachodzi korelacja między wiekiem uczestników wypraw, a najwyższą wysokością, jaką udaje im się osiągnąć.

Ponieważ zarówno dane dotyczące wieku uczestników, jak i wysokości, zawierają braki, usuwamy te przypadki. Po usunięciu obserwacji zawierających braki, zostaje do przeanalizowania 7492 przypadków.

Korelacja między wiekiem, a najwyższą wysokością, którą osiągnęli uczestnicy jest równa  $-0.093$ , zatem nie możemy mówić o skorelowaniu tych dwóch zmiennych. Potwierdza to wykres 1. Jest to ciekawy wniosek, świadczący o tym, że to nie wiek wpływa na osiągnięcie wyższych wysokości.

#### 3.2. Sukces w kolejnych dekadach

W celu zbadania ilości sukcesów w kolejnych dekadach, tworzymy dodatkową kolumnę, w której jest zapisana informacja o dekadzie, na podstawie kolumny `year`.

W danych brakuje informacji na temat wypraw w latach 40-tych, dlatego ta dekada zostaje pominięta. Może to być spowodowane II Wojną Światową, która odbywała się w tych latach. Są to jednak jedynie przypuszczenia.

W pierwszej kolejności na wykresie 2 przedstawiamy boxploty najwyższej osiągananej wysokości w poszczególnych dekadach.

Następnie zliczamy wszystkie sukcesy w każdej dekadzie. Wykres 3 ilustruje sumę sukcesów dla kolejnych dekad.

Na wykresie 2 zauważyć możemy, że wraz z upływem dekad, wspinacze osiągają coraz wyższe wysokości. Dostrzegamy również tendencję wzrostową mediany, oznacza to, że coraz więcej osób zbliża się do osiągnięcia sukcesu.

Wzrost sukcesów obserwujemy również na wykresie 3. Wraz z upływem lat pojawiają się wartości odstające, spowodowane mogą być wzrostem popularności wspinaczki wysokogórskiej, rozwojem technologii oraz wprowadzeniem wypraw komercyjnych. Przez co wspinaczka na Mount Everest stała się dostępna również dla mniej wprawionych wspinaczy, którzy kończyli swoją przygodę na niskich wysokościach.

#### 3.3. Najwyższa osiągnana wysokość w zależności od pory roku

Przeanalizujemy teraz ilość ekspedycji w zależności od pory roku.

Najbardziej popularną porą roku, którą wybierają ludzie do wspinaczki jest wiosna, co potwierdza wykres 4. Nie jest to zaskoczeniem, ponieważ panują wtedy najbardziej dogodne warunki atmosferyczne.

Analogicznie na wykresie 5 zauważamy, że wiosną ludzie docierają zdecydowanie wyżej niż na przykład zimą, podczas której warunki nie są sprzyjające. Zatem dochodzimy do wniosku, że wybór pory roku w znaczącym stopniu wpływa na sukces całej wyprawy.

### 3.4. Analiza sukcesu w zależności od płci i użycia tlenu

Kolejną częścią analizy jest zbadanie zależności sukcesu od płci i użycia tlenu. W tym celu generujemy wykres mozaikowy.

Na podstawie wykresu 6 obserwujemy, że wspinacze, którzy odnieśli sukces, w znacznej większości wspomagali się tlenem. Jak możemy zauważyć, kobiet odnoszących sukces jest zdecydowanie mniej niż mężczyzn, ale szczegółowa analiza zależności płci znajduje się w następnym punkcie.

### 3.5. Wpływ płci na osiągnięcie sukcesu

Ostatnim krokiem jest analiza wpływu płci na odniesienie sukcesu. W tym celu sumujemy wszystkie obserwacje, w których odniesiono sukces. Następnie normalizujemy wynik, porównując do ilości wypraw dla każdej płci.

Na podstawie wykresu 7 zauważamy, że więcej kobiet osiąga sukces, jednak jest to nieznaczna różnica, dlatego nie można wyciągnąć jednoznacznego wniosku czy płeć ma znaczący wpływ. Ciekawe spostrzeżenie jakie nasuwa się analizując ten wykres, to fakt, że niezależnie od płci sukces osiągany jest w mniej niż połowie przypadków.

## 4. Wnioski i podsumowanie

Celem naszego raportu było przeanalizowanie jakie czynniki wpływają na odniesienie sukcesu, czyli zdobycie szczytu Mount Everest.

Analiza doprowadziła nas do następujących wniosków: płeć oraz wiek wspinacza nie są wyznacznikami sukcesu. Potwierdził to znikomy poziom korelacji pomiędzy wiekiem, a najwyższą osiąganą wysokością. Do tej pory można było także zauważyć nieznaczającą różnicę w ilości sukcesów kobiet i mężczyzn, na rzecz kobiet. Jest więc to ciekawe zagadnienie do kolejnych badań, które dotyczyłyby predyspozycji każdej z płci do tego sportu.

Natomiast szanse na osiągnięcie sukcesu i zdobycie szczytu są z dekady na dekadę coraz większe. Podejrzewać możemy, że jest to zasługa czynników zewnętrznych takich jak postęp technologiczny i łatwa dostępność do informacji, dzięki którym można odpowiednio przygotować się do wyprawy.

Okazuje się również, że wybór pory roku ma kluczowe znaczenie. Najlepszą porą roku do wspinaczki wysokogórskiej okazuje się wiosna. Mimo, że niektórzy wspinacze wybierają również jesień, to najwyższe wysokości i największa szansa na zbliżenie się do szczytu osiągnięta jest właśnie wiosną.

Kolejnym czynnikiem, który ma wpływ na odniesienie sukcesu jest korzystanie z tlenu podczas wspinaczki. Wspinacze, którzy odnieśli sukces, w znacznej większości wspomagali się tlenem.

Podsumowując, z każdym rokiem szansa na zdobycie szczytu zwiększa się. Jeśli dodatkowo zdecydujemy się korzystać z tlenu i wybierzemy wspinaczkę wiosną przybliżamy się do sukcesu.

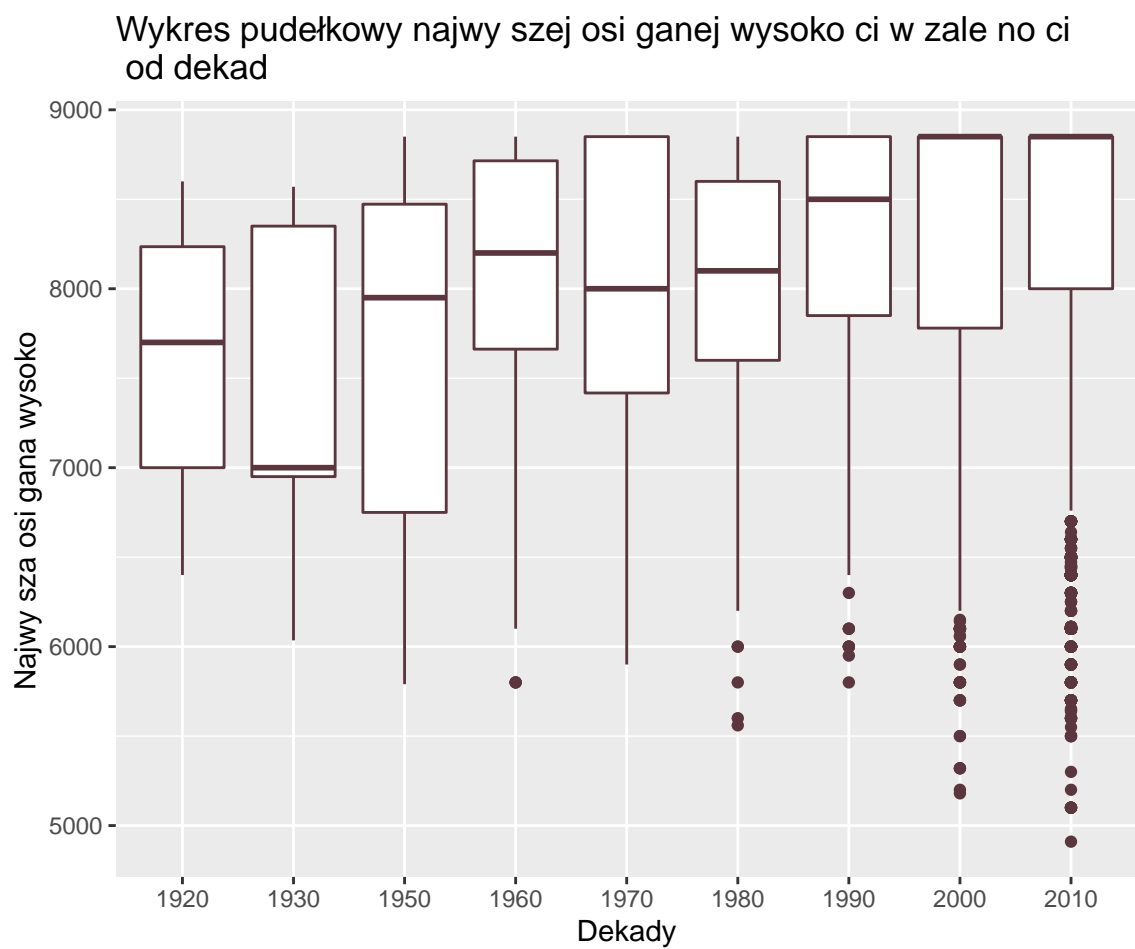
Temat ten nie został jeszcze wyczerpany, oprócz analizy predyspozycji płci do wspinaczki wysokogórskiej, można byłoby rozszerzyć raport na inne szczyty Himalajów i porównać wyniki.

## 5. Wykresy

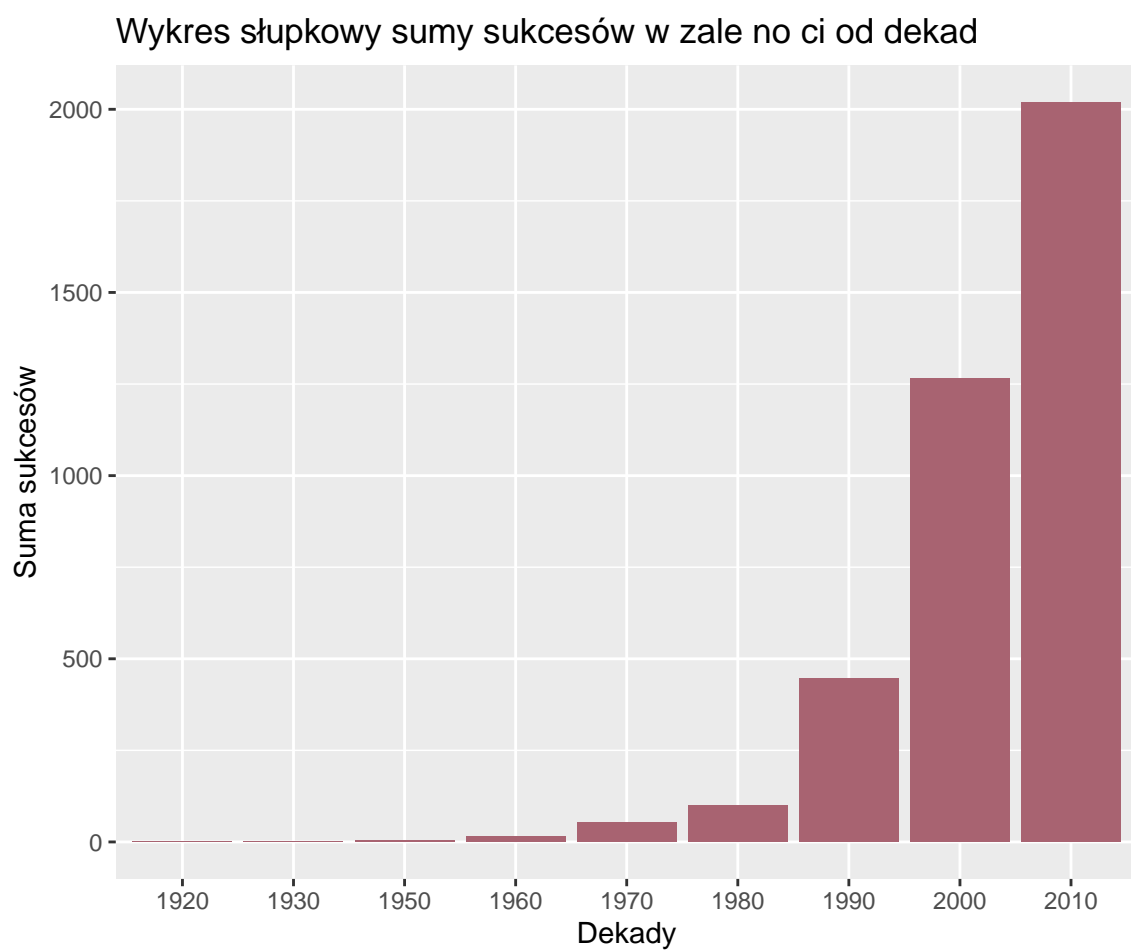
Wykres najwyższej osiągniętej wysokości w zależności od wieku uczestników



Rysunek 1. Wykres najwyższej osiągniętej wysokości w zależności od wieku uczestników.

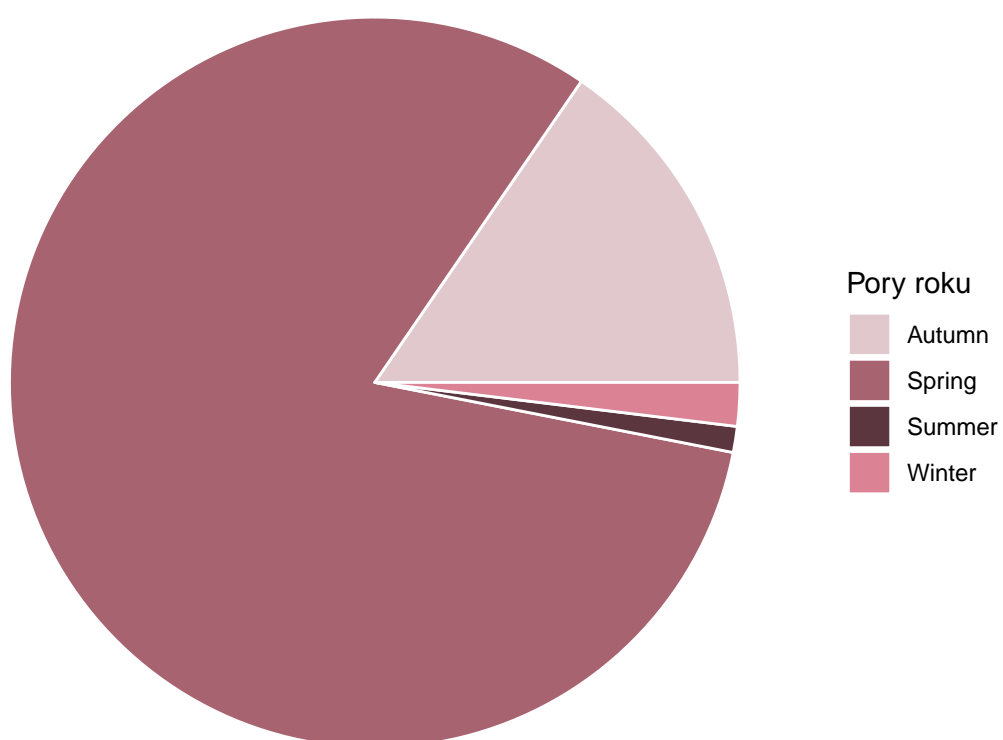


Rysunek 2. Wykres pudełkowy najwyższej osiągniętej wysokości w zależności od dekad.

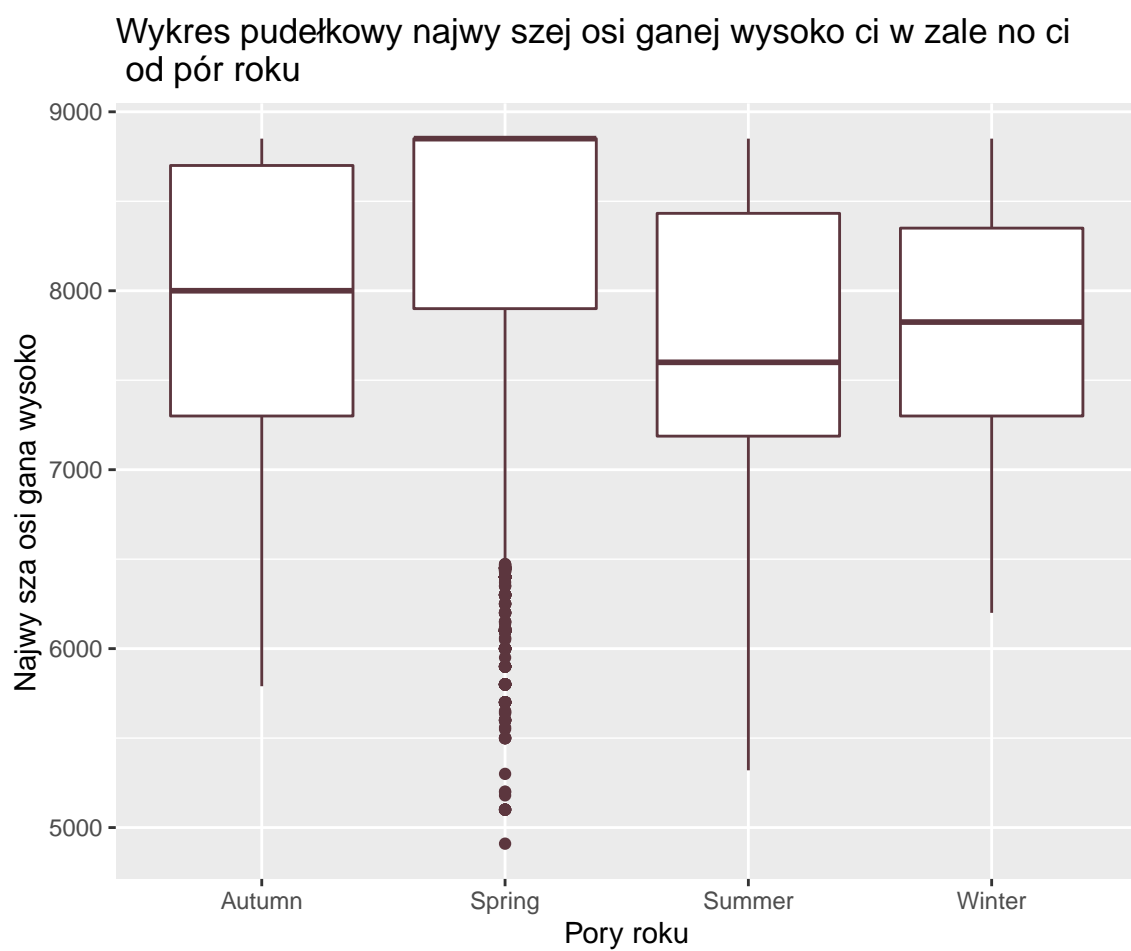


Rysunek 3. Wykres słupkowy sumy sukcesów w zależności od dekad.

Wykres kołowy przedstawiający ilość wypraw w różnych porach roku



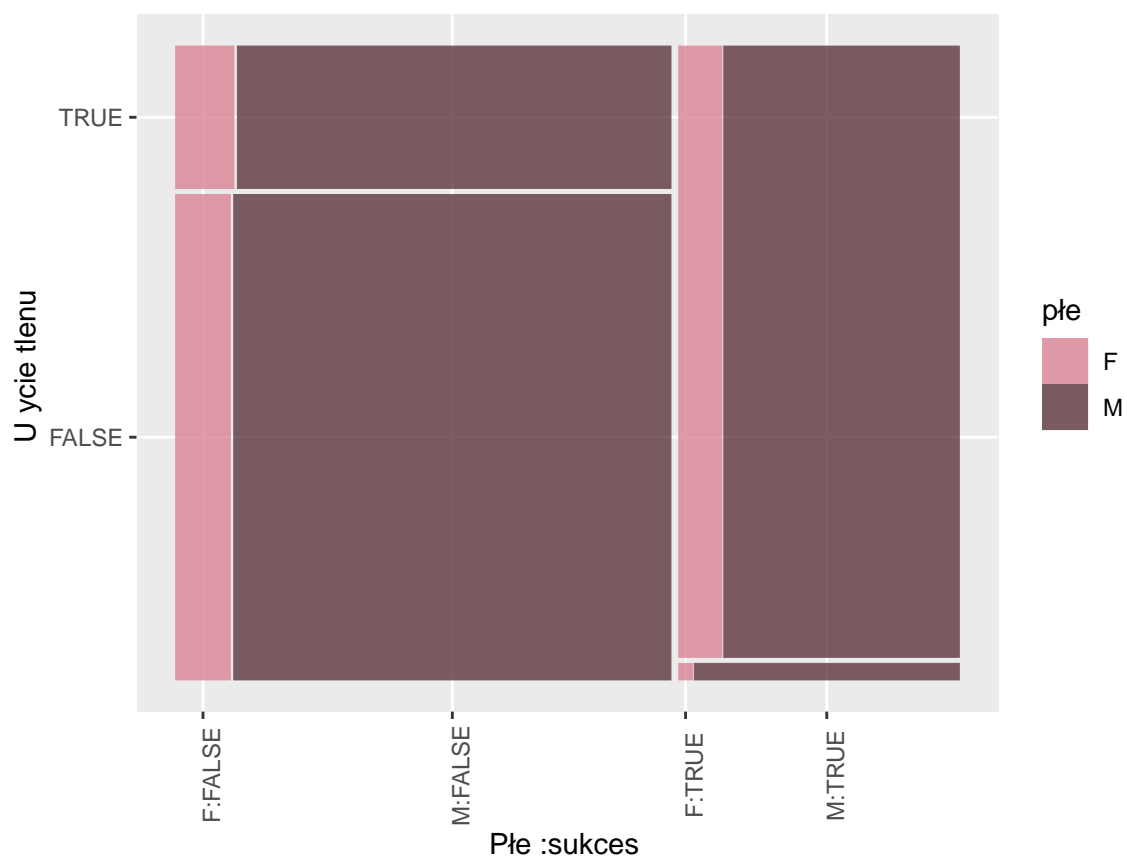
Rysunek 4. Wykres kołowy przedstawiający ilość wypraw w różnych porach roku.



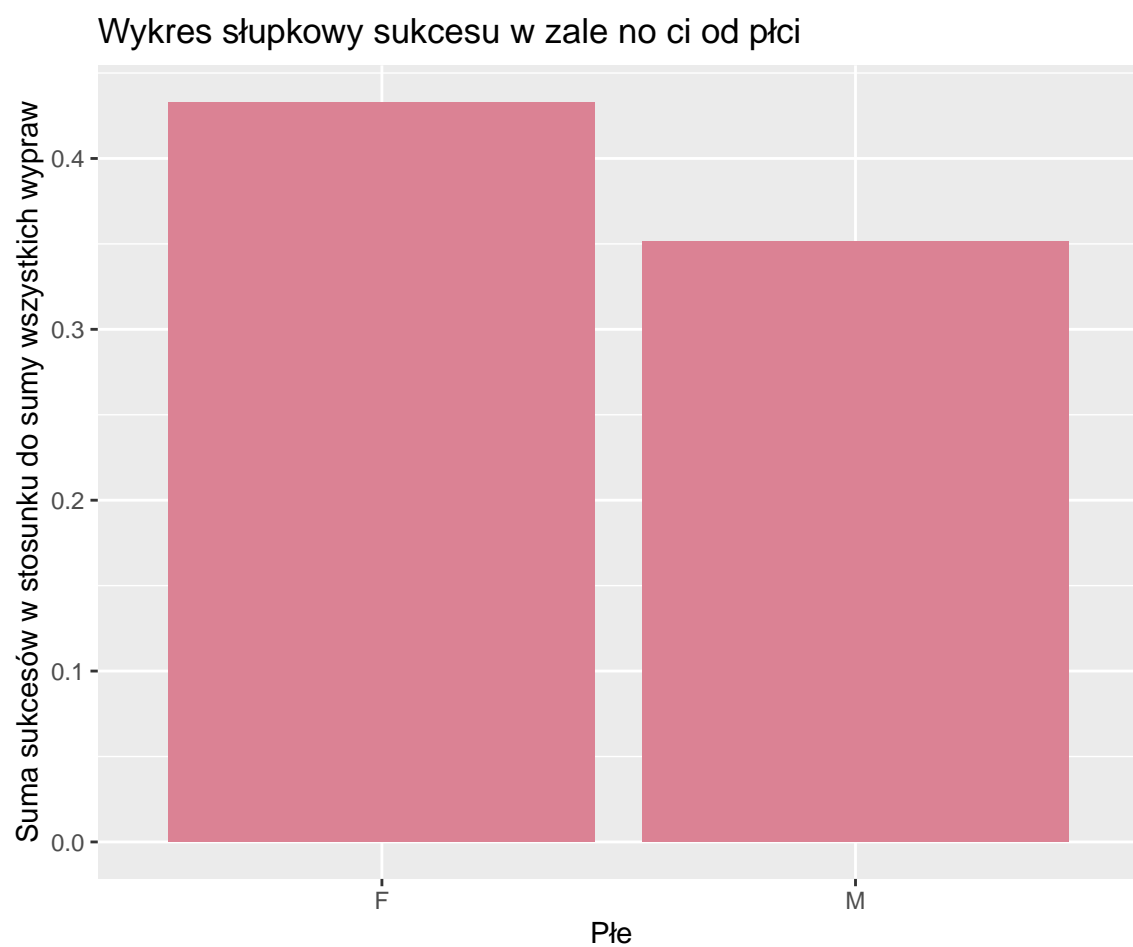
Rysunek 5. Wykres pudełkowy najwyższej osiągniętej wysokości w zależności od pory roku.



Wykres mozaikowy zależności sukcesu od płci i użycia tlenu



Rysunek 6. Wykres mozaikowy zależności sukcesu od płci i użycia tlenu.



Rysunek 7. Wykres słupkowy sukcesu w zależności od płci.