

A 2016-os magyar teljesítménytúrák adatainak elemzése

Ferencsi Tamás, tamas.ferenci@medstat.hu

2017. február 04.

Bevezető gondolatok és számítástechnikai megjegyzések

A magyar teljesítménytúrák adatainak elemzése nem nehéz feladat, ugyanis a Teljesítménytúrázó Társasága (<http://www.teljesitmenyturazoktarsasaga.hu/>) van olyan kedves és minden évben egy – számítógéppel is jól feldolgozható – Excel-táblában összefoglalja a túrák összes fontos adatát. Ezek közül most hármat fogok felhasználni: a távot, a szintet (össz-szintemelkedés) és a túra tájegységét (hegységet).

Az egyetlen „számítástechnikai” nehézség a túra tájegységének meghatározása: sajnos az Excel-tábla (nem tudom milyen oknál fogva, de ilyen szempontból nagyon peches módon) egyetlen dolgot nem tartalmaz, a tájegységet. A probléma megoldását az tette lehetővé, hogy az online túranaptár viszont megadja minden túrához a hegységet, és szerencsére az ott szereplő túranév pontosan ugyanaz, mint ami az Excel-táblában is megjelenik. Ilyen módon, számítógéppel letöltve a túranaptárt, össze tudtam kapcsolni a túrákat a tájegységekkel (is).

A távnál és a szintnél csak azokat az értékeket fogadtam el, melyek értelmes számot tartalmaztak, így azok a túrák, ahol például az intervallum jellegű megadás („20-25 km”) vagy pontatlan érték („kb. 15 km”) szerepelt, nem jelennek meg a mostani elemzésekben. A tájegységek kapcsán egy definíciós problémába futottam: bizonyos túrákhoz több hegység is fel van sorolva, mert a túra útvonala mindent érinti, így adódik a kérdés, hogy ezeket a túrákat melyik hegységhez soroljam az elemzésekben...? Jobb híján úgy döntöttem, hogy ilyen esetekben a túrát *mindegyik* tájegységhez hozzárendelem. (Így egy túra többször is megjelenik az ilyen elemzésekben; annyiszor, ahány hegységet érint.) Ennek a módszernek nyilvánvaló hátránya ez a többszöröződés, ám jobbat nem tudtam tenni.

Az elemzésből kizártam azokat a túrákat, amelyek nem voltak gyalogosan teljesíthetőek (a jellegben nem szerepelt a „Gy” betű), amelyekhez nem volt tájegység, táv vagy szint (ez utóbbi kettő a fenti értelemben) megadva, amelyiknél a megadott táv 0 méter volt, és azokat, amelyek külföldön zajlottak, akár csak részben is, tehát külföldi helyszín¹ is szerepelt a tájegységek között.

Az elemzés R statisztikai környezet alatt készült, a vizualizáció `lattice` csomaggal, a jelentés pedig `knitr`-rel, `RMarkdown` formátumban. A kódot úgy írtam meg, hogy a lehető legáltalánosabb legyen (így például könnyen alkalmazható legyen más évekre is). A teljes forráskódot, és a reprodukálható jelentést közzétettem `GitHub`-on, a következő repositoryban: https://github.com/tamas-ferenci/TTT_TuraStatisztika.

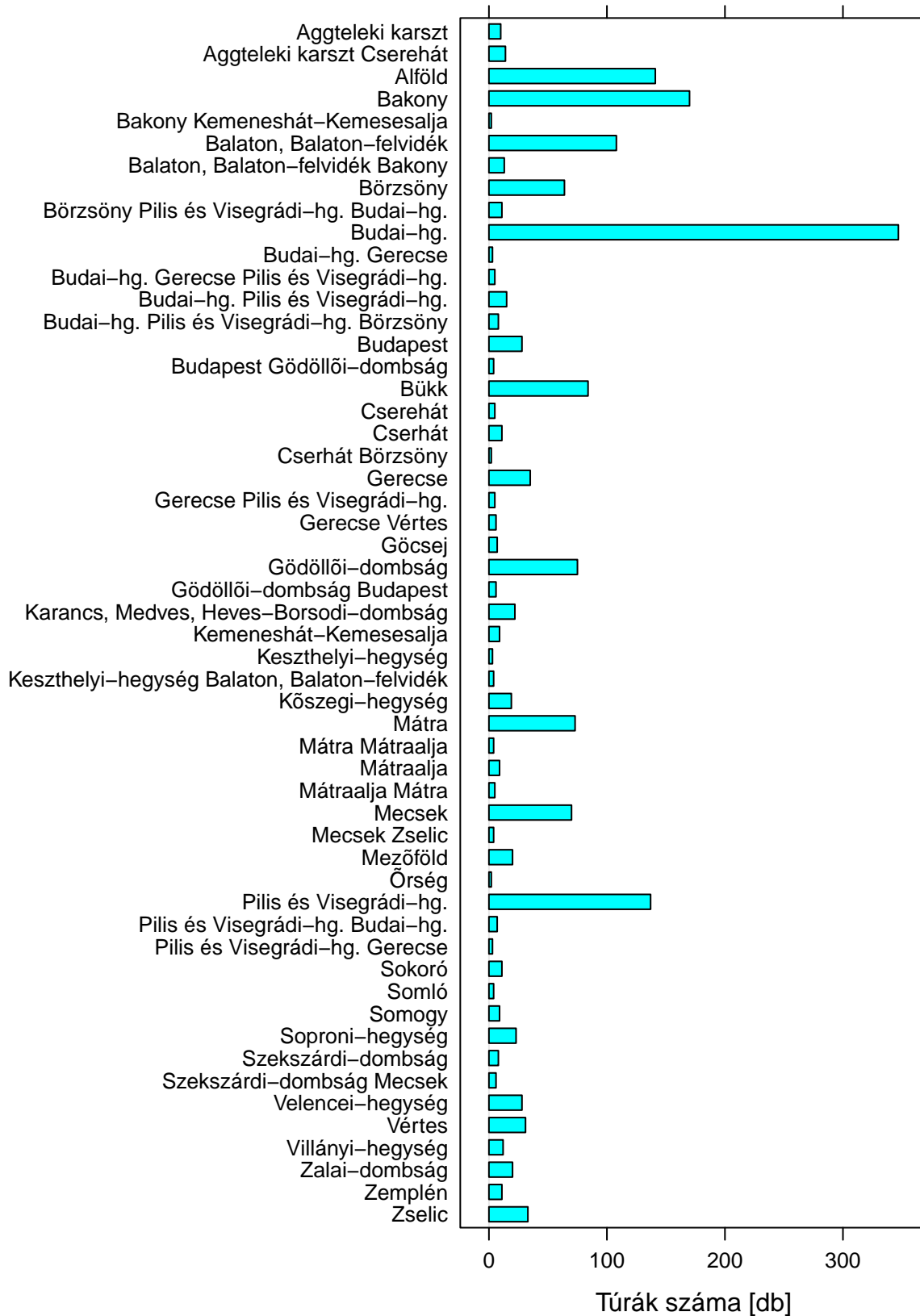
A jelentés elérhető PDF formátumban a következő címen: https://github.com/tamas-ferenci/TTT_TuraStatisztika/blob/master/TTT_TuraStatisztika_Report.pdf (letölthető a jobb oldalt felül lévő Download feliratú gombbal).

Táv, szint- és hegységadatok egyváltozós elemzése

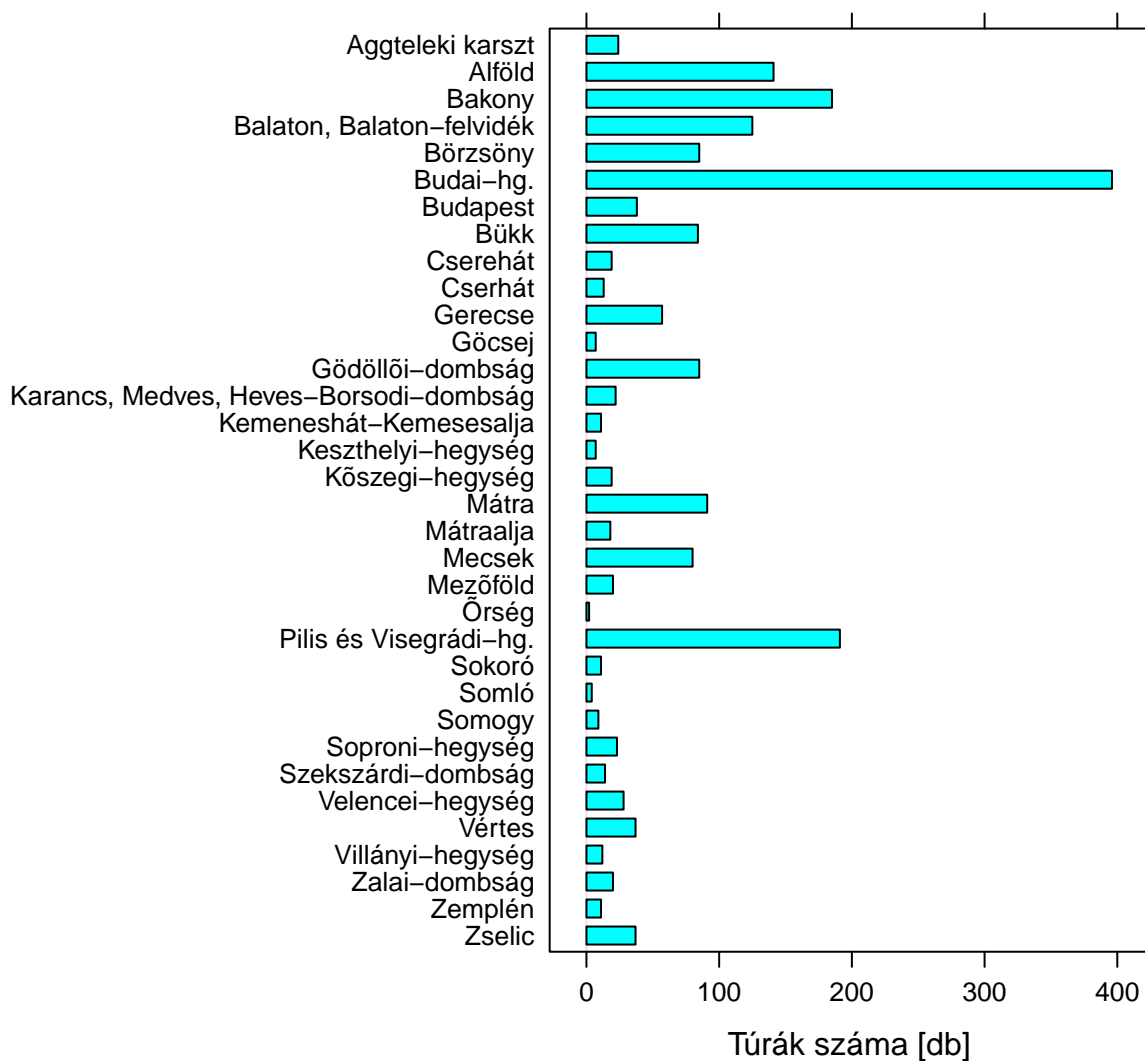
Az adatbázis eredetileg 2452 túrát tartalmaz. Ebből a nem gyalogos túrákat kizárva marad 2166, a hegység nélkülieket és külföldieket kizárva 1780, táv vagy szint nélkülieket és 0 távúakat kizárva pedig 1766.

A túrák megoszlását tájegység szerint – egyelőre még a bevezetőben írt összevonás nélkül – a következő ábra mutatja:

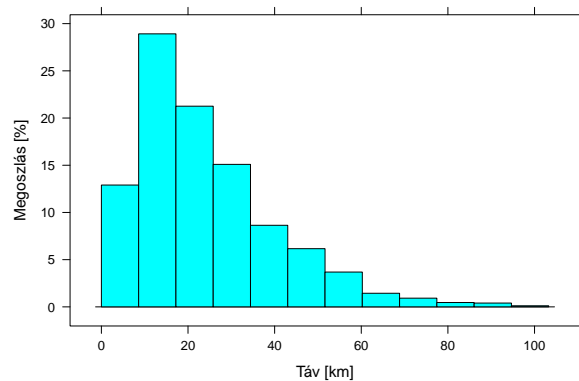
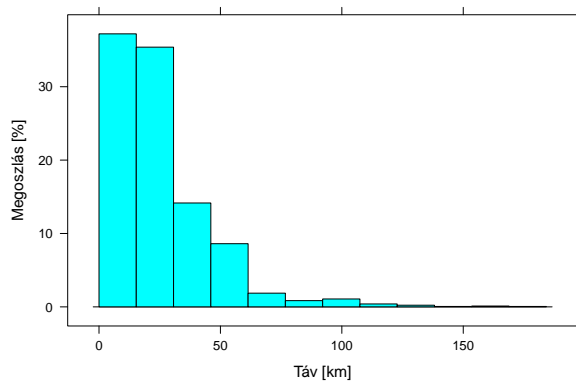
¹A következő tájegységek voltak a külföldiek: Ausztria, Bihar-hegység, Csehország, Erdély, Horvátország, Lengyelország, Németország, Románia, Szerbia, Szlovákia, Szlovénia, Ukrajna (Kárpátalja).



Jól látszik a rengeteg, több tájon áthaladó túra (és ebből adódóan a nagyon kis számok). Áttekinthetőbb képet kapunk a bevezetőben vázolt összevonás után:



A túrák távjainak eloszlását a következő ábra mutatja (hisztogramon). Mivel van egy-két extrém hosszú túra, melyek miatt a legtöbb túrát tartalmazó régió elég kicsire össze lenne nyomva, így a jobb oldalon külön látható a 100 km-nél rövidebb túrákra az eloszlás:



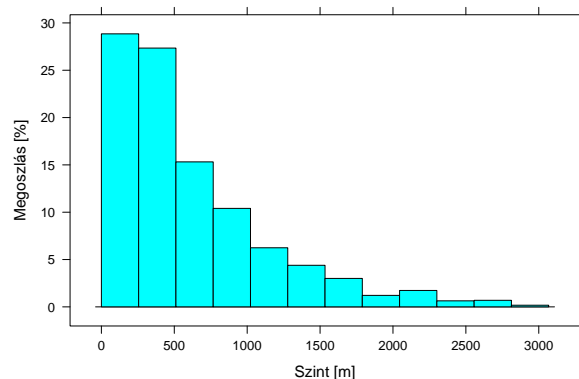
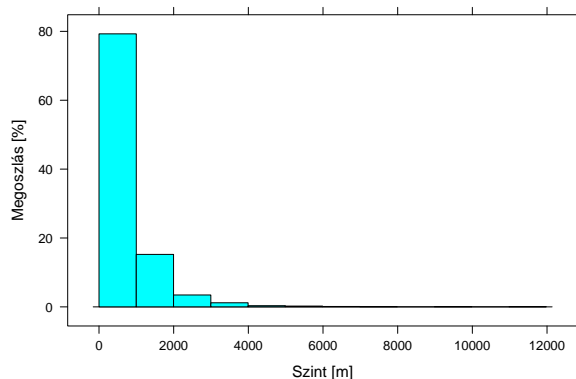
Konkréten kiemelve, számszerűen a legrövidebb túrák:

Túra neve	Dátum	Táv [km]	Szint [m]	Mered. [‰]	Hegység
Óbudavár 0,1 Éjjel	03. 19.	0,1	5	50	Balaton,
					Balaton-felvidék
Óbudavár 0,1 - A	05. 08.	0,1	5	50	Balaton,
legeslegrövidebb túra					Balaton-felvidék
Decathlon Mátra Túra -	04. 02.	1	0	0	Bakony
Törp túra 1					Mátra
Kálvária Kör	04. 02.	1,16	120	103	Villányi-hegység
Óbudavár 1	05. 08.	1,5	40	27	Balaton,
					Balaton-felvidék
					Bakony

És a leghosszabbak:

Túra neve	Dátum	Táv [km]	Szint [m]	Mered. [‰]	Hegység
Mecseki Mátrix 130	04. 16.	135	3570	26	Mecsek
Galyavár 140	11. 25.	140,22	9120	65	Mátra
3x50	10. 21.	153,9	4996	32	Budai-hg. Pilis
					és Visegrádi-hg.
Vadrózsa 160	09. 10.	165,48	6277	38	Mátraalja Mátra
Galyavár 110 mérföld (177	11. 25.	177,12	11520	65	Mátra
km)					

A túrák szintjeinek eloszlása, hasonlóan jobb oldalon csak a 3000 méternél kisebb szintemelkedésű túrák eloszlása ábrázolva:



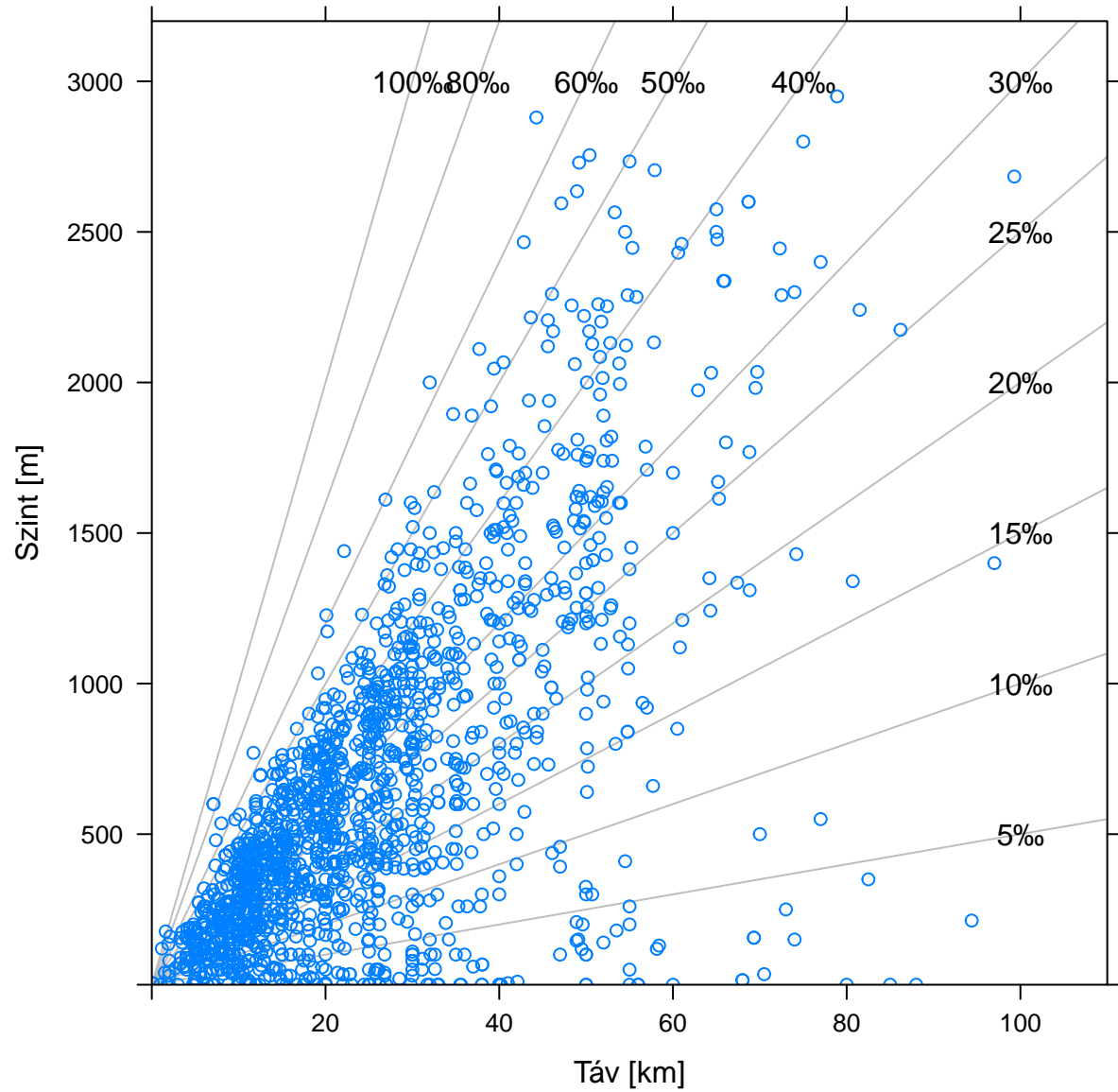
A legnagyobb szintemelkedésű túrák (a legalacsonyabbakat nincs értelme kitáblázni, hiszen több tucat van, aminél ez nulla):

Túra neve	Dátum	Táv [km]	Szint [m]	Mered. [‰]	Hegység
Mátra 115	06. 04.	121,02	6020	50	Mátra
Vadrózsa 160	09. 10.	165,48	6277	38	Mátraalja
					Mátra
Galyavár 110	11. 25.	110,7	7200	65	Mátra
Galyavár 140	11. 25.	140,22	9120	65	Mátra
Galyavár 110 mérföld (177 km)	11. 25.	177,12	11520	65	Mátra

Táv- és szintadatok együttes elemzése

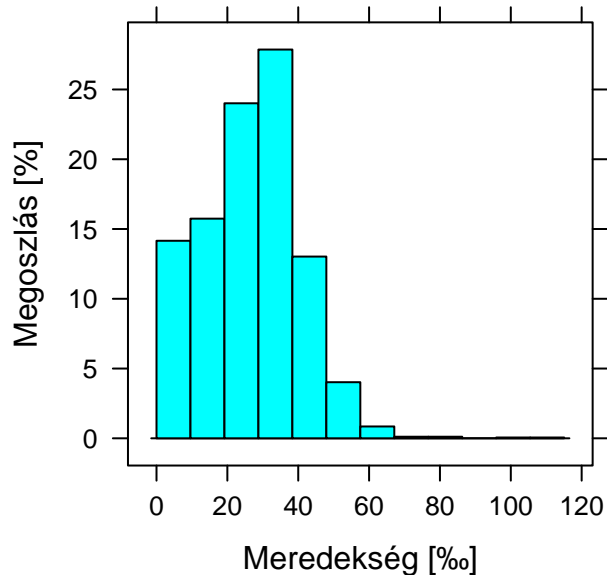
Érdekes kérdés annak vizsgálata, hogy a táv és a szintemelkedés hogyan oszlik meg, ha *együtt* vizsgáljuk őket, nem külön-külön, tehát ha egyszerre vesszük figyelembe, hogy mennyi egy túra távja és szintje. Annál is inkább, mert ezek nagyon természetesen adnak egy új paramétert (figyelmes szemlélő már az előbbi táblázatokban is észrevehette őket): a túra meredekségét, ami a szint és a táv hányadosa. (Ha az előbbi méterben mérjük, utóbbit kilométerben, akkor a meredekséget ezrelékben (‰) kapjuk meg.) Ez ilyen formában egyfajta „átlagos” meredekség, ami messze nem tökéletes mutató, hiszen a valóságban az is számít ugyancsak, hogy az emelkedések hogyan oszlanak meg a túra útvonalán, de ezekből az adatokból ez a legjobb, amit ki tudunk számolni.

A táv- és szintadatok egy kézenfekvő ábrázolása a szóródási diagram: minden pont egy túra, a vízszintes koordinátája a távja, a függőleges a szintje. Íme ez a 2016-os magyar teljesítménytúra-naptár (az áttekinthetőség kedvéért a korábban használt 100 km-es és 3000 méteres szűkítéssel):



Az ábra egy további extrát is tartalmaz: mivel a szint és a táv meghatározza a meredekséget, így ez a fenti síkban is feltüntethető. Ez az ábra meg is teszi ezt: halvány vonalak jelzik az adott meredekséghez tartozó táv–szint kombinációkat (a végén feltüntetve a meredekség számszerű értékét is). Ilyen módon az ábrán nem csak a táv és a szint, de a meredekség is érzékelhető.

A rend kedvéért azért nézzük meg a meredekségnek is az eloszlását:



A legmeredekebb túrák (a legkevésbé meredek ezúttal sem túl érdekesek):

Túra neve	Dátum	Táv [km]	Szint [m]	Mered. [‰]	Hegység
Európai sporthét 2016 - Esti sétatúra	09. 14.	2,1	160	76	Budai-hg.
Nyikom-csúcsfutás	04. 10.	7,1	600	85	Mátra
Nyikom-csúcs túra és vetélkedő	04. 10.	7,1	600	85	Mátra
Kálvária Kör	04. 02.	1,16	120	103	Villányi-hegység
Őszi Hazajáró 1,6km vezetett túra	10. 30.	1,6	177	111	Pilis és Visegrádi-hg.

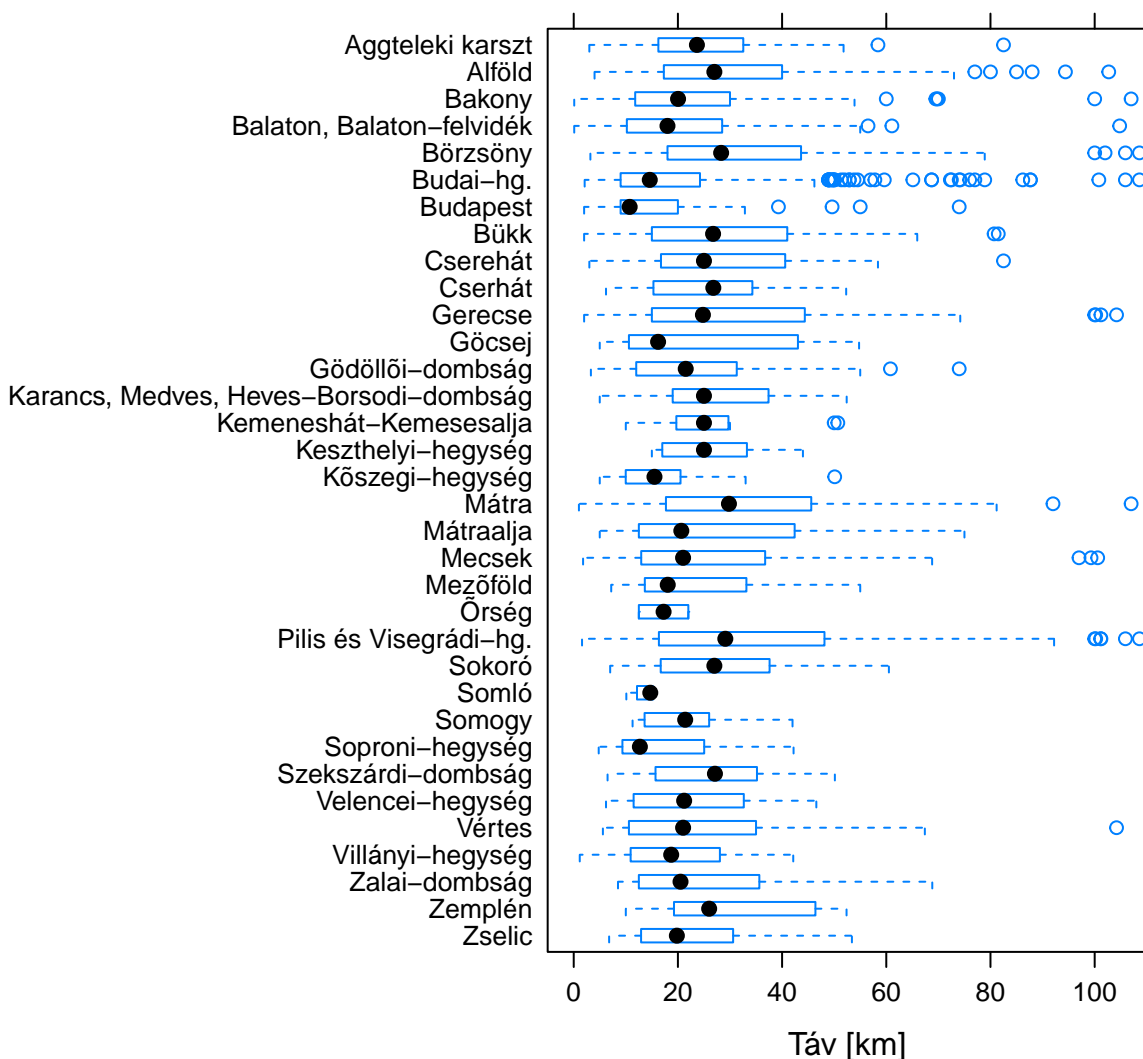
Látható, hogy ezek jórészt igen rövid túrák, így jöhetett ki az extrém meredekség, ezért érdekes lehet megnézni a legmeredekebb túrákat csak azon túrák körében, melyek legalább 10 km hosszúak:

Túra neve	Dátum	Táv [km]	Szint [m]	Mered. [‰]	Hegység
Kékes 2x	08. 06.	20,1	1227	61	Mátra
Bükk 900-as csúcsai	07. 30.	32	2000	62	Bükk
Galyavár Félmaraton	11. 26.	22,14	1440	65	Mátra
Galyavár 110 mérföld (177 km)	11. 25.	177,12	11520	65	Mátra
Galyavár 110	11. 25.	110,7	7200	65	Mátra
Galyavár 80	11. 26.	81,18	5280	65	Mátra
Galyavár Maraton	11. 26.	44,28	2880	65	Mátra
Galyavár 60	11. 26.	59,04	3840	65	Mátra
Galyavár 140	11. 25.	140,22	9120	65	Mátra
Dr. Téry Ödön nehéz teljesítménytúra	11. 19.	11,7	770	66	Pilis és Visegrádi-hg.

A tájegység bevonása az elemzésbe

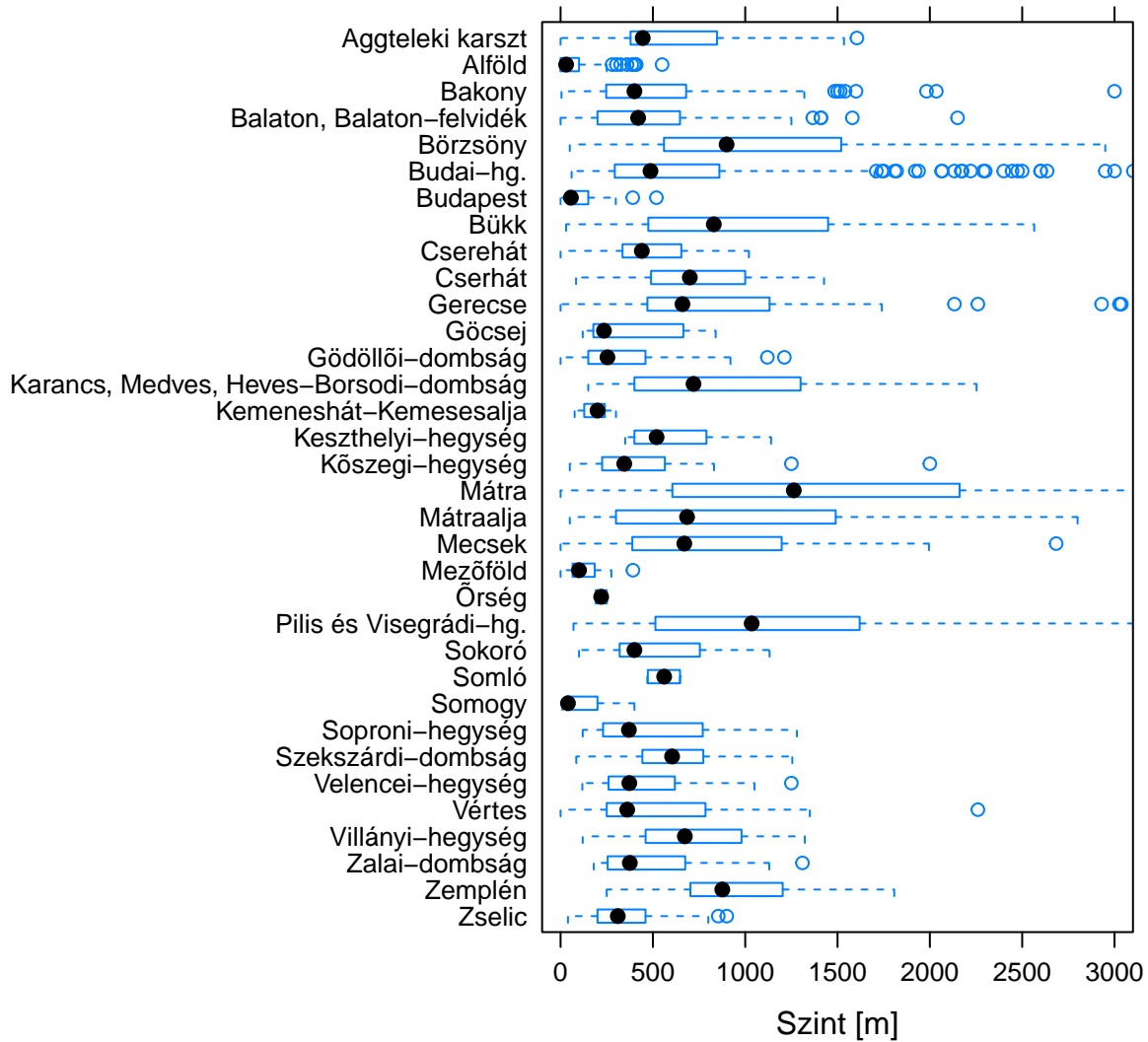
Egészítsük ki most a fenti elemzéseket a tájegységgel! A táv és szint együttes elemzésének megbontása még hegység szerint is áttekinthetetlen lenne, ezért vizsgáljuk őket külön-külön.

A teljesítménytúrák távjainak eloszlása, tájegység szerint (boxplottal² ábrázolva, ezúttal is 100 km-re szűkítve az ábrát):

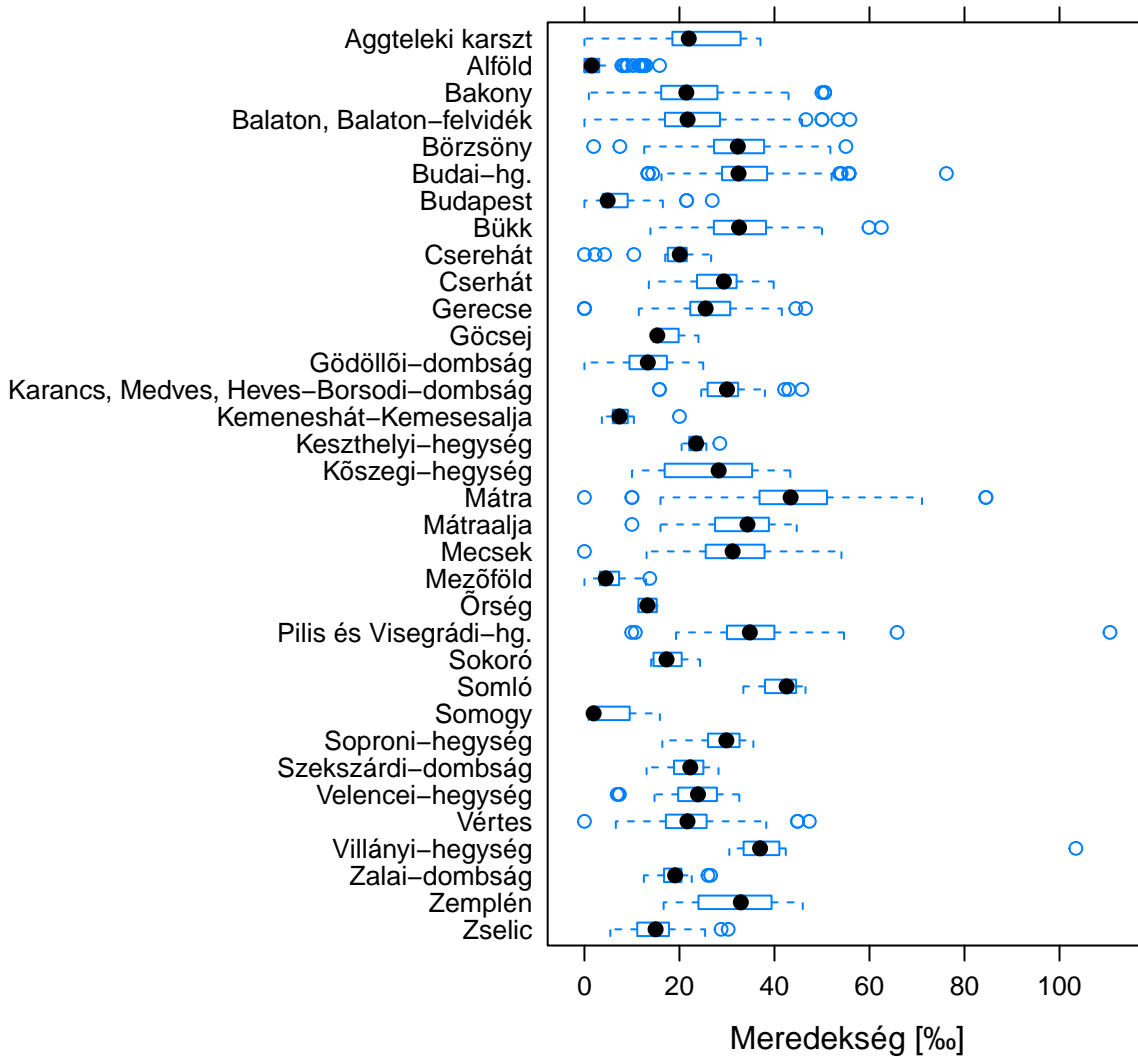


A túrák szintemelkedéseinek eloszlása boxploton, tájegység szerint (itt is 3000 méterre szűkítve az ábrát):

²A boxploton az adatok eloszlását egy – közepén pöttyöt tartalmazó – doboz, és az abból kinyúló antennák adják meg. Az értelmezéséhez azt kell tudni, hogy a pötty az adatok felezőpontjánál van (tehát annál az értéknél, amire igaz, hogy az adatok fele kisebb nála, fele nagyobb), ezt szokás mediánnak nevezni, a doboz alja és teteje pedig az alsó és felső negyedelőpontnál (aminél az adatok negyede kisebb és háromnegyede nagyobb, illetve háromnegyede kisebb, negyede nagyobb), ezt hívják alsó és felső kvartilisnek. Például a Mátránál a doboz alja 17,8 km-nél, a teteje 46,8 km-nél, a pötty pedig 29,9 km-nél van; ez azt jelenti, hogy a mátrai túrák negyede rövidebb 17,8 km-nél, háromnegyedük hosszabb ennél, a túrák fele rövidebb 29,9 km-nél, felük hosszabb, végül háromnegyedük rövidebb 46,8 km-nél, és csak negyedük hosszabb ennél is. Ilyen módon a boxplot jól érzékelhetővé teszi, hogy milyenek egy adott hegységben az „átlagos” hosszak (minél jobbrább van a pötty, annál hosszabbak a túrák), és azt is, hogy ebben mennyi a szóródás (minél szélesebb a doboz, annál nagyobb). A dobozból kinyúló antennák a nagyon kilógó értékek nélküli minimumot és maximumot jelzik, az antennákon túli egyes karikák pedig ezeket a nagyon eltérő értéket mutató túrákat (a „nagyon kilógás” egy bevett statisztikai definícióját használva).



A meredekségek eloszlása pedig:



Mindhárom tényező együttes elemzése

Mint az előbb is említettem, a táv- és szintadatok együttes eloszlása eleve több mint ezer adatot jelent, ha ezeket még a több tucatnyi tájegység szerint is megbontjuk, az teljesen áttekinthetetlen lenne.

Éppen ezért a következőkben egy új trükköt alkalmazok: nem az egyes túrákat ábrázolom, hanem a hegységenkénti átlagokat! Azaz minden hegységre kiszámolom, hogy az adott hegységet érintő túráknak mennyi a közepes hossza és a közepes szintje, és ezeket ábrázolom (a már megszokott formában, szóródási diagramként, feltüntetve a meredekségeket is). Hogy az esetleges nagyon kilógó értékű túrák ne húzzák el az átlagot, középértékként inkább a mediánt használjuk:

